MONTHLY NOTICES

OF THE

ROYAL ASTRONOMICAL SOCIETY

Vol. 107 No. 3

MEETING OF 1947 MAY 9

Professor W. M. H. Greaves, President, in the Chair

The election by the Council of the following Fellows was duly confirmed:—

John Herbert Finch, 97 Kingsbridge Road, Morden, Surrey;
Edward Alexander Michael Grimmond, Osterley, 15 Walrond Road, Swanage, Dorset;
Albert Henry McHugh, 67 Park Lane, Wembley, Middlesex;
William Douglas Oliphant, B.Sc., 36 Herriot Row, Edinburgh 3; and
Eric Richard Osler, Lamorna, Carisbrooke Road, Hucclecote, Gloucester.

The following candidates were proposed for election as Fellows of the Society, the names of the proposers from personal knowledge being appended:—

Michael Percy Dandridge, 13 Pickhurst Rise, West Wickham, Kent (proposed by C. A. Ronan);
Herbert Fletcher, M.Sc., 7 Salt Hill Way, Slough, Bucks (proposed by E. C. Bullard); and
Albert George Shrimpton, M.Sc., B.Sc. (Eng.), F.P.S.L., 16 Kingsway, Wembley, Middlesex (proposed by Sir H. Spencer Jones).

Seventy-three presents were announced as having been received since the last meeting.

Professor M. G. J. Minnaert delivered the George Darwin Lecture, taking as his subject "The Fraunhofer Lines of the Solar Spectrum" (see page 274).

After the lecture the President presented the Gold Medal to Professor Minnaert, to whom it had been awarded for his outstanding contribution to solar physics in general and solar spectrophotometry in particular; he spoke as follows:

"It is now my very pleasant duty to present the Society's Gold Medal to Professor Minnaert for his outstanding contributions to Solar Physics in general and Solar Spectrophotometry in particular. This is one of the occasions on which our Medallist is also our George Darwin Lecturer, and that being the case it

is not necessary for me to deliver a long address setting forth the grounds of the award; you have had the privilege of hearing our Medallist survey some of his own work, and you will need no justification from me of the decision of the Council to honour Professor Minnaert. Nevertheless, it is my privilege to place before you on this occasion a few considerations about our Medallist and lecturer.

"In the first place we have met to-day to do honour to a pioneer. Professor Minnaert's name is associated with the development of astronomical spectrophotometry, a comparatively young subject which has proved to be a most potent weapon of research. He was one of the first to apply the new methods to Solar Physics, and over many years his contributions have been highly original and valuable. On the observational side he has been responsible for the development of instrumental technique and for its application to a wide variety of problems. He was one of the first, working in collaboration with his colleague Pannekoek, to obtain absolute values for the intensities of chromospheric lines; we recall in this connection his photometry of the flash spectrum on the occasion of the total solar eclipse of 1927 June 29 which yielded a valuable list of intensities expressed in ergs per second per unit solid angle. Other researches by him include studies of the corona and of the brightness of the Sun near the limb and photometric investigations of the surface features of Venus and the Moon. But I must make special reference to his researches on the Fraunhofer Lines. He is primarily, I think, an observational astronomer but the progress of astronomy does not depend on observations alone but also on the application of human intelligence to their interpretation, that is on concomitant theoretical researches; and the name of Minnaert, together with those of his colleagues Mulders and Slob, have taken an honoured place in the literature of our science for their discussions of the relation between the total intensity of a Fraunhofer line and the number of absorbing atoms per square centimetre of the Sun's photosphere.

"The advancement of Astronomy depends in part on individual contributions of outstanding originality by eminent astronomers, and in the ranks of such we find our Medallist. But it also depends on the capacity of men of merit to apply themselves to the direction of a routine programme, often based on the results of previous individual research, the aim of which is to make available systematic information which must form the foundation of future assaults on the secrets of Nature. We glory in the fact that such tasks have never been shirked, and that the energies of the most outstanding names in Astronomy have been devoted to the obtaining of the observational information collected in our catalogues of star places, radial velocities, magnitudes, parallaxes and other indispensable data. Let not anyone presume to say that these energies have in any way been wasted or misapplied, we know that this work has simply got to be done as well as it can be done, and that this can only be so if the capabilities of the best of us are applied to it. It is thus fitting that the energies of our Medallist should have been devoted to the direction of the preparation of the first Photometric Atlas of the Solar Spectrum, which gives us the solar intensities, expressed in units of the continuous spectrum for all wave-lengths from $\lambda 3332$ to $\lambda 8771$. For the preparation of the Utrecht atlas special photometric apparatus was required which was devised by Professor Minnaert and his colleague Dr Houtgast, and its completion and publication marks a new epoch in the history of Solar Spectroscopy. The Utrecht atlas is the result of international cooperation and several scientific institutions have been concerned

in its preparation. In particular, the value of the atlas depends to a considerable extent on the high quality of spectrograms at Mount Wilson. But, nevertheless, the bulk of the actual work was performed by Professor Minnaert and by his colleagues Mulders and Houtgast, and so it can be regarded as one of the valuable contributions to astronomical knowledge which have originated from Utrecht, where our Medallist leads a team of astronomers who flourish under his valued directorship.

"To-day we honour our Medallist. We honour him as an outstanding astronomer and it is the privilege of many of us to honour him as a friend. To this I will add that we in this country, which escaped the horrors of enemy occupation in the long years of the recent war, honour him and his colleagues in Holland for the stand they made for the freedom of the human mind and spirit at a time when it must have seemed to them that the torch of freedom had gone out for a very long time. In particular, we honour our Medallist who, when he was an inmate of a hostage camp, endured much at the hands of those to whom freedom of thought, a heritage which is especially valued by anyone concerned with the advancement of knowledge, represented an enemy which they, in their brutal folly, vainly attempted to crush.

"Professor Minnaert; in handing you the Gold Medal of our Society, which represents the highest honour within our power to bestow, I am expressing the feeling of every Fellow when I say that not only does this Medal signify the very high opinion which we have formed of your work in the past, but it is also a symbol of the confidence with which we look forward to the future activities of yourself and of the good team that you lead".

NOTE ON NON-STATIC STELLAR MODELS

F. Hoyle and R. A. Lyttleton

(Received 1947 January 6)

Summary

The non-static cases that arise in models involving discontinuities of mean molecular weight are discussed. It is shown that a recent discussion by Ledoux is unnecessarily restrictive and that additional possibilities must be considered.

---

1. The present note is written in reply to a recent interesting paper* by Ledoux on "Stellar models with convection and with discontinuity of the mean molecular weight", which is a continuation of work begun by the present writers † in 1942. A number of additional questions are considered, but before proceeding to a discussion of these matters it may be useful to begin by mentioning the agreement that seems to have been reached over the contents of HL I.

The main topic discussed in HL I was the effect of a discontinuity of mean molecular weight occurring in the atmosphere of a star. If μ_i, X_i, n_i denote the mean molecular weight, the fraction by mass in the form of hydrogen, and the polytropic index, all for the inner region, and μ_e, X_e, n_e are the corresponding quantities in the external region then a condition

$$\mu_i(1-X_i^2)/\mu_e(1-X_e^2)=(n_e+1)/(n_i+1), \qquad (1)$$

where $$\mu_i=2/(1+3X_i), \qquad \mu_e=2/(1+3X_e)$$

must be satisfied at the interface in order that the radiative transfer be continuous. The relation (1) applies to the case in which

(*a*) μ_i, X_i are constants in the inner region, and μ_e, X_e are constants in the outer region;

(*b*) the stellar material consists entirely of hydrogen and the heavy elements producing the photoelectric opacity.‡

The importance of (*b*) will be seen from the discussion given below in Section 2.

The condition (1) for the continuity of radiative transfer is not always consistent with the assumption of static equilibrium for the whole star. Thus an upper limit can be derived for the right-hand side of (1) if we are given that the star is in static equilibrium. For in this case the value of (n_i+1) cannot be appreciably § less than $\gamma/(\gamma-1)$, where γ is the effective ratio of specific heats (the case where radiation pressure is not negligible can be included by using the "effective" ratio

* P. Ledoux, *Ap. J.*, **105**, 305, 1947.

† F. Hoyle and R. A. Lyttleton, *M.N.*, **102**, 218, 1942. This paper will be referred to as HL I, and a second paper, at present in course of publication, will be referred to as HL II.

‡ Although one of the present writers has recently pointed out (*M.N.*, **106**, 255, 1946) that opacity by electron scattering is likely to be more important than was formerly believed, photoelectric opacity is used in the present paper in order to facilitate comparison with Ledoux's work.

§ It might be thought that (n_i+1) could become appreciably less than $\gamma/(\gamma-1)$ if the interface occurred at the surface of the convective core of the star. This, however, is not the case, for as Cowling has shown, *M.N.*, **96**, 51, 1935, the convective transport of energy would then become prohibitively large.

of specific heats instead of the material ratio of specific heats), whereas the radiative equation

$$\frac{4}{3}ac\,T_e^3\frac{dT_e}{dr} = -\frac{\alpha\rho_e^2(1-X_e^2)}{T_e^{3\cdot5}}\cdot\frac{L}{4\pi r^2} \qquad (2)$$

for the outer region requires $(n_e+1)\leqslant 4\cdot25$ for stability at the surface of the star. Accordingly the largest permissible value of $(n_e+1)/(n_i+1)$ consistent with static equilibrium is $4\cdot25(\gamma-1)/\gamma$. But values of X_e and X_i can be chosen so that the left-hand side of (1) exceeds $4\cdot25(\gamma-1)/\gamma$. In such cases it follows, since the radiative transfer certainly cannot be discontinuous, that we must still satisfy (1) and accept the situation that a non-static state must develop in the star. Thus we divide all possibilities into two categories. In the first category we have all cases in which a static equilibrium state for the star is consistent with (1) and the second category in which there is no static state consistent with (1).

In HL II it is pointed out that (1) requires a discontinuity in the polytropic index whenever a discontinuity in molecular weight occurs. It is satisfactory that Ledoux apparently agrees with this conclusion, since he uses (1) throughout his paper. This is in distinction to Schönberg and Chandrasekhar who use * the erroneous condition $n_i=n_e$ in discussing a discontinuity in molecular weight. Indeed it would seem that Ledoux has expressed no item of disagreement with the discussion of the static models given in HL I. Thus it is only in questions concerning non-static cases that differences of opinion arise. In order to make the position clear it is desirable to state that HL I is almost entirely concerned with static cases. Some qualitative references are made to non-static cases. In particular the following sentence occurs in the Summary to HL I:—

"The causes leading to the conditions for giant stars and non-static atmospheres being brought about are briefly discussed". Ledoux takes strong exception to this sentence and apparently on no other grounds he states that our conclusions "are without any real foundations". The work given below is not so much directed against specific features of Ledoux's analysis but is written to show that the evolution of non-static stars is by no means as certain as he maintains.

2. *The effect of helium.*—The purpose of the present section is to discuss the modification of (1) that arises when the stellar material is regarded as being composed of hydrogen, helium, and the heavy elements producing the photo-electric opacity. Thus if X is the fraction by mass in the form of hydrogen, and Y is the fraction by mass in the form of helium then (1) becomes

$$\mu_i(1-X_i-Y_i)(1+X_i)(1+n_i)=\mu_e(1-X_e-Y_e)(1+X_e)(1+n_e),$$

where $\mu_i=4/(2+Y_i+6X_i), \qquad \mu_e=4/(2+Y_e+6X_e).$ (3)

The effect of this modification is to produce an important broadening in the category of static cases. Thus cases that were static according to (1) will remain static according to (3), but a range of problems that were non-static according to (1) become static according to (3). This situation applies to the problem considered by Schönberg and Chandrasekhar † and as this particular problem is discussed in detail in Ledoux's paper it is useful to regard it as an illustrative example of the difference between (1) and (3).

The S-C problem considers a discontinuity in molecular weight that arises

* M. Schönberg and S. Chandrasekhar, *Ap. J.*, **96**, 162, 1942.
† M. Schönberg and S. Chandrasekhar, *loc. cit.*

from the conversion of hydrogen to helium within the inner region (the conversion in the outer region being negligible). This case evidently requires

$$1-X_i-Y_i=1-X_e-Y_e,$$

since the difference between μ_i and μ_e is regarded as being due to the effect of thermonuclear reactions and not to an inherent difference in the fraction by mass in the form of heavy elements. A further simplification arises because $X_i=0$ in the S-C problem, and it is reasonable to simplify the matter still further by putting $Y_e=0$. This means that the star is regarded for the present purposes as composed almost entirely of hydrogen and heavy elements at the time of its condensation from the interstellar gases. The hydrogen is converted into helium within the convective core and no large scale mixing takes place between the core and the outer envelope (if such a mixing process took place it would be pointless to consider the S-C problem). Thus in the limiting case considered in the S-C problem, where the hydrogen in the core is regarded as being completely exhausted, we have the following very simple conditions

$$\left.\begin{aligned} &1-X_i-Y_i=1-X_e-Y_e,\\ &X_i=Y_e=0. \end{aligned}\right\} \tag{4}$$

Under these conditions it is easy to show that the S-C problem will belong to the static category if

$$2(1+3X_e)/(2+X_e)(1+X_e)\leqslant 4{\cdot}25(\gamma-1)/\gamma. \tag{5}$$

It follows from this inequality that when radiation pressure is negligible and $\gamma=5/3$ the S-C problem always belongs to the static category; for the left-hand side of (5) has a maximum just extending $4/3$ for $0<X_e<1$, whereas the right-hand side of (5) is equal to $17/10$ when $\gamma=5/3$.

The adoption of (3) in place of (1) does not affect in principle the considerations of HL I, since static models were the main topic of discussion in that paper.* But the present result does affect the numerical work given in the later part of Ledoux's paper, because he considers the S-C problem with $\gamma=5/3$ to belong to the non-static category. Thus although the numerical values following Table I of his paper must be nearer to the correct solution than the values given originally by Schönberg and Chandrasekhar, the new values still do not represent a proper solution of the S-C problem.

The remarks following (5) show that non-static cases will occur if

$$4{\cdot}25(\gamma-1)/\gamma<4/3. \tag{6}$$

If we regard γ as being less than $5/3$, owing to the effect of radiation pressure, it is seen from (6) that non-static cases must occur when $\gamma<51/35$. It is estimated that this would require $\beta<0{\cdot}65$ and for $\mu_i=4/3$ this requires a star with mass greater than $10_{\odot}$. Thus stars of fairly large mass that have exhausted the supply of hydrogen in the inner region can fall into the non-static category. This means that although the category of non-static cases is narrowed by the adoption of (3) in place of (1) it is not removed.

* The broadening of the category of static cases also allows more favourable " red giant " cases to be discussed. Thus the limitation in radius for the examples given on page 222 of HL I arose from the limitation imposed by (1) on μ_i/μ_e. The condition (3) allows cases with much larger values of μ_i/μ_e, and this leads to far bigger radii. Details of these considerations will be published as soon as integrations that take account of the mass in the atmosphere are available.

3. The purpose of the present section is to discuss a model for the non-static problem; that is, in addition to the possibility considered by Ledoux. Now that the effect of helium has been demonstrated it is proposed to revert to (1) with $\gamma = 5/3$, since this greatly facilitates comparison with Ledoux's paper.

The continuity of the radiative transfer requires one of the three following cases to occur at the interface in the non-static problem:—

$$\left.\begin{array}{lll} \text{(i)} & n_i < 1{\cdot}5, & n_e = 3{\cdot}25 \\ \text{(ii)} & n_i = 1{\cdot}5, & n_e > 3{\cdot}25 \\ \text{(iii)} & n_i < 1{\cdot}5, & n_e > 3{\cdot}25 \end{array}\right\} \text{ at the interface.}$$

The third case can be regarded as a combination of the other two, and need not be discussed separately. Ledoux considers only case (i) and he gives no reasons for rejecting (ii). Thus the conclusions derived in Sections 2, 3, 4 of his paper are no surprise, since they are *put into* the analysis as soon as his equation (18) is assumed. By this assumption Ledoux excludes all possibility of a non-static outer region and thereby avoids discussion of the question at issue. The present purpose is to consider case (ii).

The material of the star will be regarded as being initially static. Non-static states will be introduced at a later stage. At the interface $n_i = 1{\cdot}5$ and n_e is adjusted to make the radiative transfer continuous. The distribution of density ρ_e and temperature T_e are chosen so that the equation of mechanical equilibrium

$$\frac{\mathfrak{R}}{\mu_e}\frac{d}{dr}(\rho_e T_e) = -\frac{GM(r)\rho_e}{r^2} \tag{7}$$

is satisfied. If we consider only distributions in which n_e is finite and greater than 1·5 then it can be shown from (7), taken together with the definition of n_e, *viz.*

$$n_e \frac{dT_e}{dr} = \frac{T_e}{\rho_e}\frac{d\rho_e}{dr}, \tag{8}$$

that ρ_e, T_e tend simultaneously to zero at the surface of the star. There are infinitely many distributions satisfying these conditions.

This freedom in the distribution of ρ_e, T_e arises because no condition of radiative equilibrium has been imposed on the outer region. The transfer of energy by radiation, L_{rad} say, is given by

$$L_{\text{rad}} = -\frac{16acr^2T_e^{6\cdot5}}{3\alpha\rho_e^2(1-X_e^2)}\cdot\frac{dT_e}{dr}, \tag{9}$$

and by introducing the polytropic index this can be written, in virtue of (7), in the form

$$L_{\text{rad}} = \frac{16\pi acG\mu_e}{3\mathfrak{R}\alpha(1-X_e^2)}\cdot\frac{MT_e^{6\cdot5}}{\rho_e^2(1+n_e)}. \tag{10}$$

Equation (10) shows that L_{rad} is finite at the surface for all distributions in which n_e tends to 3·25 at the surface. None of these distributions, however, gives $dL_{\text{rad}}/dr = 0$ in the outer region, for, by hypothesis, we are dealing with non-static cases, whereas if this condition were satisfied the requirements for a static state would be fulfilled. But the breakdown of static conditions will not become appreciable until a time has elapsed of order

$$\text{(Total heat content of the outer region)}/L,$$

and for the examples considered in HL I this time is greater than 1000 years

even for massive stars (that is, for stars with large L). Thus the breakdown of static equilibrium in the distributions discussed above is a *second order* effect, not to be compared with a *first order* violation of (7), which would produce an appreciable change in ρ_e, T_e in a time of a few hours.

This discussion suggests that it is worth while investigating the effect of (7) not being satisfied in *second order*. The type of motion that the writers have in mind is an adiabatic oscillation of the star, the oscillation being mainly confined to the outer region. Although such an oscillation is only a second order effect so far as (7) is concerned it can have an important effect on the mean rate of radiative transfer. Thus the mean rate of transfer $(L_{\text{rad}})_m$ say is

$$\frac{16\pi acG\mu_e}{3\Re\alpha(1-X_e^2)}\left[\frac{MT_e^{6\cdot5}}{\rho_e^2(1+n_e)}\right]_{\text{mean}}, \tag{11}$$

which differs from the rate of transfer in the static case by terms that depend on the square of the amplitude of the oscillation. The work of Rosseland * suggests that these terms become important when $\delta R/R \simeq 0{\cdot}25$, where R is the radius of the star and δR is the amplitude of the oscillation at the surface.

The present suggestion is that the pulsation can be arranged so that

$$d(L_{\text{rad}})_m/dr = 0.$$

The problem would be to find a mean distribution of ρ_e, T_e and an amplitude of pulsation such that this is the case. Thus it was pointed out above that there are a great variety of distributions of ρ_e, T_e satisfying the condition that the rate of transfer at the interface shall be continuous and also such that ρ_e, T_e tend simultaneously to zero at the surface of the star. These distributions all satisfy (7) and consequently are possible *mean* values of the density and temperature for the pulsation case.

The difficulty of carrying out this programme can be realized from Rosseland's investigation. Thus it is already difficult to work out the terms in $(\delta R/R)^2$ when the mean values of ρ_e, T_e are *specified*, but the present problem is still more difficult because the mean values of ρ_e, T_e have to be inferred from the terms in $(\delta R/R)^2$. The only feasible method for solving the problem would appear to be by using large calculating machines in a trial and error method. It is recognized that the above remarks will remain no more than a suggestion until this has been done. Indeed the writers would have been content to leave the matter in the cautiously worded terms used in HL I, if Ledoux's assertion that our conclusions are "without any real foundations" had not made it desirable to discuss the question rather more fully. The present object is to show that case (ii) must be considered on its merits, and that Ledoux is not justified in simply ignoring it.

4. It is interesting to compare cases (i) and (ii) of Section 3. In the former case the outer region remains static and a non-static state develops in the inner region. The writers' belief is that this is another example of "the tail wagging the dog" but we have not succeeded in rigorously proving it. Indeed when instability problems arise where there is more than one possible line of evolution, it is usually extremely difficult to give a conclusive theoretical proof of the way the system must develop. The difficulties that arise in the present problem are well instanced by the discussion following equation (50) in Ledoux's paper. In comparing case (i) with a static model he finds a reduction in the radius of the star by about three parts in ten

* S. Rosseland, *M.N.*, **103**, 233, 1943.

thousand and he concludes by saying that " for all practical purposes the radius can be regarded as unchanged ". The question is what are " all practical purposes " ? A reduction of radius by three parts in ten thousand is sufficient to supply enough gravitational energy to start the pulsations described above. The velocity so acquired by the material would exceed one per cent of the velocity of escape from the surface of the star, and this is already a velocity of several km. per sec. This matter is raised, not with the purpose of criticizing the specific arguments put forward by Ledoux, but to show that very small effects that might normally be regarded as negligible must be taken into account in the present problem.

Although it is not desired to enter in detail on the complicated arguments that arise in trying to decide between cases (i) and (ii), there is one particular question that seems worth mentioning. Ledoux has calculated the time required for a mixing process to bring case (i) back to a static state, and he finds the time to be so short that he concludes that the mixing is practically instantaneous. This seems to the writers to be a misleading way of stating the matter, for the time that is of significance in Ledoux's case is the time required to build up the convective region in which $n_i < 1{\cdot}5$. Thus it seems almost certain that, starting from a static state, the time required to build up Ledoux's case will be long compared with the period of the adiabatic oscillations mentioned in Section 3. Accordingly if oscillations can occur with the required properties then it is to be expected that they will take up the departure from static conditions long before Ledoux's convective region is built up.

In view of the theoretical stalemate discussed above it seems permissible to refer to observations for a resolution of the difficulty. According to the work of HL I the radius of a star with a discontinuity of molecular weight occurring in the outer parts becomes large compared with the main sequence value (the mass being the same) when μ_i/μ_e is close to 2. The investigation of Section 2 shows that this can occur without a non-static condition necessarily developing. Thus we can have static giant stars of low surface temperature. But if radiation pressure becomes important a stage can be reached where an entirely static state is no longer possible. The case (ii) of Section 3 suggests that when this occurs the extended parts of the star will develop a free adiabatic oscillation that builds up until terms in $(\delta R/R)^2$ become important. It is expected that this will occur when $\delta R/R \simeq 0{\cdot}25$. Hence on the view proposed by the writers some red giant stars should be static and others pulsating, which is in satisfactory agreement with the observational data. On the other hand the view put forward by Ledoux, while permitting the occurrence of red giant stars, does not lead to an explanation of why many red giant stars are observed to pulsate. Although it is clear that the advantage rests with the suggestion of the present paper, the writers realize that an empirical appeal to observation is not always a satisfactory method of settling a theoretical difficulty. The position may fairly be stated as follows :—If case (ii) is correct then we obtain what seems to be a satisfactory explanation of the origin of the pulsations of the long period variables (and possibly also of cluster-type Cepheids), whereas if case (i) should prove correct then we still have to find a source of the pulsations shown by many red giant stars.

5. The analysis of the external layers given in Sections 2, 3, 4 of Ledoux's paper proceeds on the probably valid assumption that when a region occurs in which $n < 1{\cdot}5$ the transport of energy by convection is negligible compared with the transfer by radiation. But this assumption is certainly not valid when the

discontinuity of molecular weight occurs near the convective core of the star. For this reason equation (61) of Ledoux's paper is open to serious question and the statement following this equation that "the situation is therefore very similar to what we described under analogous conditions in the stellar envelope...." cannot be accepted without further discussion. Indeed it would seem that Ledoux realizes the dubious nature of equation (61) for he makes no use of it. Thus instead of attempting to treat the non-static problem he assumes that mixing will take place across the interface. The situation is that the work leading to (66) cannot be applied to the non-static problem until this assumption has been proved. Ledoux in fact proceeds as if there were no possibility other than the mixing process, but as seen in Section 3 it is by no means certain that this model is correct. Accordingly it is possible that the work of Section 5 of his paper does not refer at all to the evolution of a non-static star, and simply represents the consideration of a case belonging to the static category. It is clear therefore that predictions concerning the evolution of a star that are based on this work must be treated with caution.

St. John's College,
Cambridge :
1947 *January* 4.

NOTE ON THE ORIGIN OF WHITE DWARFS

F. Hoyle

(Received 1947 January 10)

Summary

A theory of the origin of white dwarfs proposed by Eddington in 1939 is examined. It is shown that the difficulties encountered by Eddington in attempting to explain the synthesis of the heavy elements from hydrogen arise from his refusal to admit the Stoner-Anderson formula for the pressure of relativistically degenerate electrons. If Eddington's theory is correct then we arrive at the following conclusions :—

(i) The material of the Galaxy was originally composed of pure hydrogen.

(ii) The majority of white dwarfs are composed largely of hydrogen.

(iii) The Fermi form of the β-decay interaction applies to the proton-proton reaction.

On the other hand if Eddington's theory is wrong then the following conclusions are reached :—

(i) All white dwarfs arise from nova and supernova nuclei.

(ii) Either the supply of white dwarfs is greater by a factor of order 100 than is suggested by the accepted statistics on the frequency of novae and supernovae, or the density of white dwarfs in the neighbourhood of the Sun is much larger than the average density for the Galaxy as a whole.

(iii) The high hydrogen content of Sirius B is due to accretion of interstellar hydrogen occurring after the formation of this white dwarf.

1. The observational data concerning Sirius B indicate that this white dwarf contains about sixty per cent by mass in the form of hydrogen. This conclusion is independent of the controversy between Eddington and Chandrasekhar on the formula for the pressure of degenerate electrons at very high densities.* Indeed the result seems inescapable, unless either the Einstein red shift of about 20 km. per sec. or the mass of $0{\cdot}95\,\odot$ is called into serious question.

The high hydrogen content of Sirius B has been discussed by Marshak † who finds difficulties in explaining the *present* structure of the star. In a recent paper Schatzman ‡ has attempted to overcome these difficulties by considering suitably chosen shell models. But the question raised by Marshak (concerning the hydrogen content of Sirius B) is of a minor character compared with the remarkable argument given by Eddington § in 1939. This argument shows that the outstanding problem is to find the sequence of stellar evolution that can lead to the formation of white dwarf stars with high hydrogen content. The difficulties raised by this question are of a far deeper significance than those that arise in attempting to explain the present properties of the star.

* Thus Eddington in *M.N.*, **99**, 595, 1939 obtains 68 per cent in the form of hydrogen, while Chandrasekhar in *Introduction to the Study of Steller Structure* gives 52 per cent.

† R. E. Marshak, *Ap. J.*, **92**, 321, 1940.

‡ E. Schatzman, *Ann. D'Astrophys.*, **8**, 143, 1945.

§ *Loc. cit.*

Eddington's conclusions are easily understood by considering the evolution that the Sun would undergo if thermonuclear reactions were not taking place in the interior. Then the continued radiation at the surface would lead to a slow collapse. The time taken for the radius to be reduced to a tenth of its present value would be about 10^8 years, but further contraction would proceed more slowly owing to the increasing importance of electron degeneracy. The pressure due to the degenerate electrons would prevent the radius being reduced below about one-fortieth of the present value (this being comparable with the radius of Sirius B). When this stage was reached radiation at the surface would lead not to further collapse but to a cooling of the interior. Such a cooling process has the property of being comparatively rapid at first, but slows up as the inner regions cool down owing to a marked decrease in the rate of radiation at the surface.*

During the stages of collapse before electron degeneracy becomes important the central temperature must rise. Eddington estimates that the central temperature would attain 10^8 deg. C. This is probably a lower limit, and we may adopt 2×10^8 deg. C. as a more likely value.† If we now reinstate the thermonuclear reactions we see that it is impossible for the Sun (with its present hydrogen content) to contract to a white dwarf state, because the rate of energy generation by the carbon-nitrogen cycle when operating at a temperature of 2×10^8 deg. C. would be enormously greater than the luminosity. Thus the assumed situation in which the Sun is contracted to a tenth of its present radius cannot be permanent. A rapid expansion would evidently occur and the radius would soon return to its present value. This conclusion raises the important question:—How was Sirius B formed?

Eddington in attempting to answer this question finds that the initial carbon and nitrogen content of the material of Sirius B cannot be greater than‡ about one part in 10^{16}. We can regard this result as being equivalent to saying that carbon and nitrogen must be wholly absent. Furthermore if carbon and nitrogen are wholly absent it is reasonable to suppose that other heavy elements are also absent. Eddington accepts the latter conclusion and regards Sirius B as being initially composed of pure hydrogen.

An additional problem of energy generation, not considered by Eddington, now arises. According to Bethe and Critchfield§ the energy generation in pure hydrogen is about 2 ergs per g. per sec. when the density is 100 g. per cm.^{3} and the temperature is 2×10^7 deg. C. Now we have seen that a much higher central temperature of about 2×10^8 deg. C. must be attained if a body of solar mass is to contract down to a white dwarf state. At this temperature and the corresponding density of about 10^6 g. per cm.^{3} the Bethe-Critchfield theory gives an energy generation of about 5×10^6 ergs per g. per sec. This is vastly greater than the value required to maintain the luminosity. Thus if the Sun were contracted to a tenth of its present radius the luminosity would correspond to a mean energy loss of

* It is estimated that about 10^9 years would be required for the central temperature to fall to 10^6 deg. C.

† The collapse would not necessarily occur through a series of homologous configurations. Consequently it is difficult to estimate the exact value to which the central temperature rises. It seems unlikely, however, that a reduction in radius by a factor of more than ten would lead only to the fivefold increase of temperature given by Eddington. The value here adopted corresponds to an increase by a factor of ten.

‡ This value is an upper limit since Eddington's central temperature is a lower limit.

§ Bethe and Critchfield, *Phys. Rev.*, **54**, 248, 1938.

only 6 ergs per g. per sec. The discrepancy is somewhat reduced by noting that the value of 5×10^6 ergs per g. per sec. applies to material at the centre of the star. It is known * that the mean energy production is only about two per cent of the central value, so that the mean energy production is 10^5 ergs per g. per sec. However, the difference between this value and the energy production required to maintain the luminosity is still very large. It may be concluded that the energy generation in pure hydrogen (on the Bethe-Critchfield theory) is too large to admit of the formation of white dwarfs from pure hydrogen by a factor of order 10^4.

This result is sufficient to destroy Eddington's argument if we accept the Bethe-Critchfield theory. But there is an alternative to this theory, which depends on accepting the form of the β-decay interaction given originally by Fermi, instead of the spin selection rule later proposed by Gamow and Teller. The Fermi interaction leads to a much lower energy production in pure hydrogen than the Gamow-Teller case. Some significance may possibly be attached to the fact that this reduction at a temperature of 2×10^8 deg. C. is by a factor of about 2×10^4, which is just of the magnitude required by Eddington's theory. This factor arises from a term (Mean thermal velocity of the protons/Velocity of light)^2, that appears in the Fermi but not in the Gamow-Teller form.†

In the past, astrophysicists, in their search for sources of energy generation, have tended to support the Gamow-Teller selection rules. But the issues involved in the present question are so important that it is desirable to emphasize that the decision between the two forms of β-decay interaction is by no means certain so far as nuclear physics is concerned. The interpretation of the often quoted He^6-Li^6 reaction is not without ambiguity, and even if it were the mechanism underlying the phenomenon of β-decay is so little understood that we cannot be sure of the validity of applying to the proton-proton reaction a result derived from a reaction involving many particle nuclei. Thus the decision between the two types of spin selection rule is still an open question, and it is permissible to adopt the Fermi form if the astrophysical data should require it. Accordingly Eddington's theory cannot necessarily be dismissed on the grounds of a too large energy production in pure hydrogen.

A minor objection to Eddington's argument may be mentioned at this stage. The observational data indicate a hydrogen content in Sirius B of about sixty per cent. It would evidently be more convenient if the hydrogen content were near a hundred per cent. The change in the Einstein red shift required to give this increase is from 20 km. per sec. to about 15 km. per sec. This change is small and may well be within the errors of observation. All re-examinations of the data seem to have been made so far with a view to reducing the hydrogen content. It would be useful to know what changes might reasonably be made for the opposite purpose of increasing the hydrogen content.

2. It is natural to consider the alternatives available to Eddington's theory. Two possibilities may be suggested:—

(1) The observational data for Sirius B are in error, and the correct values would lead to zero hydrogen content;

* F. Hoyle and R. A. Lyttleton, *M.N.*, **102**, 177, 1942.

† For further details regarding the properties of these β-interactions see Bethe and Critchfield, *loc. cit.*

(2) At the time of condensation the hydrogen content of Sirius B was zero; and the present hydrogen content of Sirius B has risen from the subsequent accretion of interstellar hydrogen.

Of these possibilities (1) seems very unplausible, but (2) evidently demands serious consideration. The attempt to explain all hydrogen observed in white dwarfs as being due to accretion taking place *after* the central temperatures have fallen to values of order 10^6 deg. C. leads to an interesting line of argument. This will now be discussed.

If we regard the hydrogen content of all white dwarfs as being zero at the time of their condensation then we must explain the sequence of stellar evolution that leads to the exhaustion of hydrogen in these stars. Two suggestions may be rejected at the outset. Firstly, it is quite unplausible to suppose that an appreciable number of stars without hydrogen can condense directly from the interstellar material, for as Dunham * has shown, the hydrogen abundance in the interstellar material is about 10,000 times greater than the abundance of the commonest metals. Secondly, the generally accepted time-scale of about 10^{10} years is far too short † for the conversion of hydrogen to helium by thermonuclear reactions to exhaust the hydrogen supply in stars with masses less than the solar mass.

It appears therefore that the only stars in which the hydrogen supply can become exhausted are stars more massive than the Sun. The mechanism whereby such stars produce white dwarfs is fairly well understood. Thus in a recent paper ‡ it was seen that stars without hydrogen enter a collapsing phase and if the mass appreciably exceeds $\odot$ the collapse continues until rotational instability arises. Material is then thrown off to infinity until the mass in the remaining nucleus becomes less than Chandrasekhar's limit §, which is close to $1{\cdot}44\odot$. This process has been associated with the nova and supernova phenomena. ‖

The process outlined in the previous paragraph undoubtedly provides a steady supply of white dwarfs. But the important question is whether the supply is sufficiently rapid. If the observed density of white dwarfs within 10 parsecs from the Sun is regarded as typical for the Galaxy as a whole, then the total number of white dwarfs in the Galaxy is of the order of 10^9. The accepted rate of one supernova per galaxy per 500 years gives 2×10^7 supernovae per galaxy in 10^{10} years, which is too small to explain the estimated number of white dwarfs by a factor of order 100. This discrepancy could be reduced if it is assumed that each supernova is disrupted into an appreciable number of fragments. ¶ However, this assumption is open to very serious question for it has been shown ** that the rate of contraction in the most rapidly collapsing star is only about one per cent of the average velocity of sound in the stellar material. Accordingly it appears most unlikely that supernovae can break into a number of separate pieces. This theoretical conclusion is in agreement with observation which indicates that material is showered off very

* T. Dunham, Jr., *Proc. Amer. phil. Soc.*, **81**, 277, 1939.

† The writer withdraws the method used for calculating the time-scale in *Proc. Camb. phil. Soc.*, **35**, 603, 1939. This method, which was based on the time required for the exhaustion of hydrogen in white dwarfs, leads to unacceptable results when applied to white dwarfs with masses appreciably less than $\odot$.

‡ F. Hoyle, "Note on Equilibrium Configurations for Rotating White Dwarfs", *M.N.*, **107**, 231, 1947.

§ Chandrasekhar's limit is $5{\cdot}75\odot/\mu_e^2$, and $\mu_e \simeq 2$ when there is no hydrogen.

‖ F. Hoyle, *M.N.*, **106**, 343, 1946.

¶ G. Gamow, *Phys. Rev.*, **59**, 617, 1939.

** F. Hoyle, *loc. cit.*

steadily in a supernova outburst. In particular, Minkowski's estimate of $15\odot$ for the mass of the *diffuse* material in the Crab nebula shows that the material thrown off by the supernova of A.D. 1054 did not consist of a series of dense "bolts" with masses of order $\odot$.

At first sight it might seem that the supply of white dwarfs from novae appreciably exceeds the supply from supernovae on account of the greater frequency of novae. But this naïve point of view ignores the result that the mass of material thrown off in a nova outburst is only of order $10^{-5}\odot$. Thus a collapsing star with mass appreciably greater than $\odot$ requires at least 10^4 outbursts of a nova type in order to reach permanent stability. This is in contrast with the supernova case where there is only one outburst. On the basis that the supply of white dwarfs is the same from novae as from supernovae, a rate of one supernova per galaxy per 500 years corresponds to about 20 novae per galaxy per year. This rate of novae is in accordance with observation. Thus it seems likely that the supply of white dwarfs from novae is of the same order as the supply from supernovae.

The above discussion suggests that there is a discrepancy between the observed number of white dwarfs and the number supplied by stars with masses appreciably greater than $\odot$. If this is the case, then neither of the possibilities mentioned at the beginning of the present section is sufficient to dispose of Eddington's theory. The alternatives presented are:—

(*a*) *Eddington's theory*:

(i) The material of the Galaxy was originally composed of pure hydrogen.
(ii) The majority of white dwarfs are composed largely of hydrogen.
(iii) The Fermi form of the β-decay interaction applies to the proton-proton reaction.

(*b*) If Eddington's theory is incorrect then

(i) All white dwarfs arise from nova and supernova nuclei.
(ii) Either the supply of white dwarfs is greater by a factor of order 100 than is suggested by the accepted statistics on the frequency of novae and supernovae, or the density of white dwarfs in the neighbourhood of the Sun is much larger than the average density for the Galaxy as a whole.
(iii) The high hydrogen content of Sirius B is due to accretion of interstellar hydrogen occurring after the formation of this white dwarf.

The decision between these two lines of evolution is evidently one of the most important problems of cosmogony. In particular (i) of (*a*) implies that the heavy elements are formed from hydrogen. The acceptance of this view has hitherto been a matter of aesthetic preference, but if Eddington's theory is correct we are compelled to accept this result.

3. In the present section we proceed on the assumption that Eddington's theory is correct. We have then to explain the formation of the heavy elements from hydrogen by processes taking place within the Galaxy. Eddington finds * a number of serious difficulties in attempting to satisfy this requirement, but it can be shown that these difficulties arise from his refusal to admit the Stoner-Anderson formula for the pressure of relativistically degenerate electrons. On the basis of this formula Chandrasekhar has shown that no *spherically symmetric* degenerate equilibrium configuration is available when the mass of a collapsing star exceeds

* Eddington, *loc. cit.*

$5{\cdot}75\odot/\mu_e^2$. For stars composed of hydrogen $\mu_e = 1$ and the limiting mass is $5{\cdot}75\odot$. When the mass exceeds $5{\cdot}75\odot$ the star must contract until either:

(1) The energy generation in pure hydrogen becomes sufficiently large to compensate for the radiation at the surface. A sufficient degree of contraction must lead to this condition being satisfied even if the Fermi form of the β-decay interaction is adopted;

or

(2) The star becomes rotationally unstable.

If the angular momentum of the star is small enough (1) must occur before (2).

It is important to notice the difference between stars with masses of order $\odot$ and stars with masses exceeding $5{\cdot}75\odot$. For small masses the contraction is limited by the occurrence of a degenerate equilibrium configuration, and if we accept the Fermi form of the β-interaction the energy generation is less than the luminosity when this stage is reached. Thus such stars form white dwarfs in the manner discussed by Eddington. On the other hand when the mass exceeds $5{\cdot}75\odot$ there is no such limit to the contraction and even on the Fermi form of the β-interaction a stage must be reached, provided the angular momentum of the star is sufficiently small, where the energy generation balances the luminosity at the surface. When this balance is reached the rate of conversion of hydrogen to helium is very rapid on account of the large luminosity that must then prevail. Indeed if no further material is added by accretion the hydrogen supply will be exhausted in a time less than 10^9 years. As soon as an appreciable quantity of helium is formed, carbon will be synthesized in triple collisions between α-particles. The star expands as the amount of carbon increases. This effect has been described by Eddington.

The above discussion shows how the hydrogen can become entirely converted into helium when the mass exceeds $5{\cdot}75\odot$. The star then follows a line of evolution that has already been described in detail.* This evolution provides for the synthesis of the heavy elements, and the eventual onset of rotational instability leads to the heavy elements being distributed in interstellar space. The crucial difference between this theory and the discussion given by Eddington is that no attempt is here made to synthesize the elements in material containing hydrogen. Thus by synthesizing the heavy elements *after* the hydrogen has been entirely converted to helium the difficulties encountered by Eddington are avoided.

The chronological history of the Galaxy according to this theory may be briefly outlined:—

(i) The Galaxy consists initially of pure hydrogen and consequently the first stars are composed of pure hydrogen.

(ii) If the mass of such a star is appreciably less than Chandrasekhar's limit a white dwarf is formed.

(iii) If the mass is appreciably greater than Chandrasekhar's limit, and the angular momentum of the star is small enough, contraction continues until the energy generation balances the luminosity. Further contraction is then arrested until the hydrogen is exhausted. When the contraction is resumed the heavy elements are synthesized and are finally distributed in interstellar space by rotational instability. The time

* F. Hoyle, *M.N.*, **106**, 343, 1946.

required for the evolution of these stars is not appreciably greater than 10^9 years.

(iv) About 10^9 years after the formation of the first massive stars the interstellar material contains appreciable quantities of heavy elements. The first trace of carbon and nitrogen is sufficient to prevent further white dwarfs being formed by direct condensation. Thus white dwarfs are formed in large numbers only during the earliest phase in the history of the Galaxy. Subsequent stellar condensations follow the main sequence. As little as one part in 10^{10} of carbon and nitrogen in the stellar material is sufficient to produce this change.

(v) The white dwarfs remain stable once their central temperatures have fallen to values of order 10^6 deg. C., even if they subsequently acquire carbon and nitrogen by accretion of interstellar material.

4. In this section it will be assumed that Eddington's theory is incorrect, but the interstellar material will still be regarded as initially composed of pure hydrogen. The difference between the present discussion and that of Section 3 is that we here examine the consequences of accepting the Gamow-Teller form instead of the Fermi form of β-interaction. In this case it is easy to show that the first stars will contract until the luminosity is balanced by energy generation in pure hydrogen. For stars with masses less than $\odot$, this stage will be reached for radii of the same order as the values occurring in the normal main sequence. But as the mass increases the equilibrium radii become appreciably less than the main-sequence values. The evolution of stars with masses greater than $5{\cdot}75\odot$ is essentially the same as in Section 3. Thus the heavy elements are synthesized and distributed in interstellar space in a similar manner. Once the interstellar carbon and nitrogen becomes appreciable subsequent stellar condensations are of the main sequence type.

St. John's College,
Cambridge :
1947 *January* 8.

THE RELATION OF COHESION TO ROCHE'S LIMIT

Harold Jeffreys

(Received 1947 February 15)

Summary

Roche's proof of the existence of a critical distance for a satellite, within which it would be broken up by tidal action, assumed the satellite fluid. It is shown that a solid satellite, or a small solid body making a casual near approach, would not be broken up even close to the surface of a planet unless its diameter was less than a critical value, which is unexpectedly large, of the order of 200 km. It appears in particular that small bodies of the sizes of the majority of the asteroids could not have been formed by close approach to Jupiter, and that fragments of ice consistent with the maximum possible thickness of Saturn's rings could not have been formed by disruption of a solid mass by tidal action.

It was proved by Roche that a small liquid satellite moving in a circular orbit about a primary, the periods of rotation and revolution being equal, would take an approximately ellipsoidal form provided that it was not too close to the primary. If, however, the mean distance was less than about twice the radius of the primary (depending somewhat on the ratio of the densities) there would be no possible permanent form and the satellite would be broken up. Many extensions of Roche's argument have been given, notably by Darwin and Jeans; the theory has been given for gaseous satellites and the restriction that the satellite is to be small has been removed. On the whole the modifications give surprisingly little change in Roche's main result. But the hypotheses all suppose the satellite to be fluid, and the corresponding problem for a solid satellite seems to have escaped explicit solution. The solution can actually be derived easily from that of the straining of an elastic sphere.

If an incompressible spherical body with density ρ, radius a, surface gravity g, and rigidity μ is deformed by a gravitational potential k_nK_n, where K_n is a solid harmonic, the stress components are given by

$$p_{ik}=-p_n\delta_{ik}K_n+\mu\left\{2(A+Br^2)\frac{\partial^2K_n}{\partial x_i\partial x_k}+(2B+C)\left(x_k\frac{\partial K_n}{\partial x_i}+x_i\frac{\partial K_n}{\partial x_k}\right)+2C\delta_{ik}K_n\right\}, \tag{1}$$

where

$$\mathrm{B}=-\frac{n+3}{2n}C\ ;\qquad A=\frac{n+2}{2(n-1)}Ca^2\ ;\qquad \epsilon_n=\frac{2n+1}{2(n-1)}Ca^n, \tag{2}$$

$$C\left\{\frac{2n^2+4n+3}{n}\mu+g\rho a\right\}=\rho k_n, \tag{3}$$

$$p_n=\mu\{(4n+2)B+2C\}+\rho\left(k_n+\frac{3g\epsilon_n}{(2n+1)a^{n-1}}\right). \tag{4}$$

The radial displacement at the surface is ϵ_nK_n/a^{n-1}, but is not required in the present problem. Gravity between parts of the small body is of minor importance in comparison with rigidity; even for a body as large as the Earth the terms in the coefficient of C are nearly as 6 to 1 for $n=2$. We shall therefore neglect terms in g.

Compressibility is also of small importance, affecting the results by only a few per cent.

Now take the body to be at distance c from the centre of a mass M and to rotate with an angular velocity ω in a plane through the radius vector. Then the deformation is derived from a gravitational potential

$$U=\frac{fM}{2c^3}(2x^2-y^2-z^2)+\tfrac{1}{6}\omega^2(x^2+y^2-2z^2). \tag{5}$$

We consider two cases. The body may be a wandering asteroid, making an accidental close approach to Jupiter or Mars. In that case it may be supposed that ω is far too small to produce disruption by itself; in some cases stress-differences due to rotation will have already been removed by plastic adjustment. In these cases we can take ω zero.

The body may however be a satellite that is being made to approach its primary by tidal friction or a resisting medium. If tidal friction in the satellite itself is strong or the satellite is triaxial like the Moon, the relation

$$\omega^2=fM/c^2 \tag{6}$$

will be maintained throughout the motion. Thus we should take

$$U=\frac{fM}{2c^3}(2x^2-y^2-z^2)\ ; \qquad U=\frac{fM}{6c^3}(7x^2-2y^2-5z^2), \tag{7}$$

according as we are considering a small body making a single approach to a larger one or a satellite gradually approaching its primary.

In both cases it is found that the greatest stress-difference is $|p_{11}-p_{33}|$ evaluated at the origin, that is, at the centre of the small body, and the respective results are

$$\frac{24}{19}\frac{fM}{c^3}\rho a^2\ ; \qquad \frac{32}{19}\frac{fM}{c^3}\rho a^2. \tag{8}$$

The different hypotheses about the rotation therefore do not affect the order of magnitude of the result.

As a specimen we consider a small satellite with the density of the Moon moving near the Earth's surface. The critical stress-differences for the rocks are about 10^9 dynes/cm.^{2}. If ρ_0 is the density of the Earth we have

$$M/c^3=\tfrac{4}{3}\pi\rho_0, \tag{9}$$

and the second of (8) gives for rupture the condition

$$\tfrac{4}{3}\pi\ .\ \frac{32}{19}\ .\ 6{\cdot}66\times10^{-8}\rho_0\rho a^2>10^9, \tag{10}$$

whence, with $\rho_0=5{\cdot}5$, $\rho=3$ g./cm.^{3},

$$a\geqslant1{\cdot}1\times10^7\text{ cm.}=110\text{ km.}$$

In other words, a satellite, and *a fortiori* an asteroid, if of rocky constitution, could graze the Earth's surface without rupture if its diameter were less than 220 km. This value is unexpectedly large, since it is well known that a column of rock a few kilometres in height would be crushed under its own weight. The difference arises from the fact that in the argument leading to equation (8) the centre of the body is free and the resultant attraction of the Earth is used up in maintaining the orbital motion. For the column of rock the resultant attraction has to be balanced by internal stresses over the base.

For the approach of a rocky body to Jupiter the result must be multiplied by about 2, since ρ_0 is then about 1·3 g./cm.^{3}. Thus if all the asteroids were once one

body, this would have been large enough to be broken by a very near approach to Jupiter; but no number of such approaches could have produced some thousands of bodies with diameters of a few kilometres.

Another interesting case is the approach of a sphere of ice to Saturn. I am informed by Mr B. B. Roberts, of the Scott Polar Research Institute, that ice cliffs usually break off at a height of about 100 feet. This would indicate a strength of the order of 3×10^6 dynes/cm.^{2}; the strength at the temperature prevailing near Saturn may be greater. Allowing for the changes in (10) we find that the critical radius for an ice sphere, if it is to be broken up by tidal action near Saturn's surface, is of the order of 3×10^6 cm. At the mean distance of the rings, of course, this would have to be increased by a factor of 3 or 4. Hence if an ice satellite ever revolved about Saturn at the mean distance of the rings, and was broken up by tidal action, its diameter was over 200 km.; and the fragments would cease to be broken up further when their diameters had been brought down to this value. Since the ring appears equally bright all round and its maximum possible thickness is given as 10 km. we can conclude that it was not formed by disruption of a solid satellite.

After the above was written, Dr M. Perutz informed me that laboratory tests on ice give an average tensile strength of 11·8 kg./cm.^{2} = $1{\cdot}16 \times 10^7$ dynes/cm.^{2}, at -5 to -10 deg. C., but remarks that individual pieces may fail at half to twice this value. In any case the estimates just given are not too low.

This investigation is supplementary to the usual theory of Roche's limit, not a substitute for it. With sufficiently large bodies it would be possible for the present criterion to be satisfied for distances outside Roche's limit. Subsequent developments would then depend on whether the material was of a type to undergo elastic failure by fracture or flow. In the former case the satellite would presumably be broken into two immediately (that is, within the time needed for a fracture wave to traverse the diameter, probably not more than a few hours, and possibly minutes). In the latter case it would adjust itself by flow towards a state of hydrostatic pressure, and ultimately the resulting deformation would make Roche's criterion applicable again; the body would be permanently distorted but not broken up.

If the present criterion for elastic failure is not satisfied there will be no disruption even if the distance is within Roche's limit. If the distance is within Roche's limit and the criterion is satisfied, the behaviour will differ according as failure is by immediate fracture or flow. For fracture, since a non-zero cohesive force is suddenly removed, the parts will separate with a non-zero difference of velocity towards the primary, and unless the rupture occurs at an apse the parts will have different energies per unit mass, and will therefore proceed in orbits with different mean distances. For plastic flow, the body will be distorted as a quasi-fluid, but will not reach a steady state since this state would itself contradict Roche's theorem. Several complications, such as the ratio of the masses of the parts and the rate of plastic adjustment, would affect the details and it does not appear possible to say without further information whether the outcome would be two independent satellites, two revolving about each other, or the equilateral triangle configuration.

St. John's College,
Cambridge:
1947 *February* 14.

THE EFFECTS OF COLLISIONS ON SATURN'S RINGS

Harold Jeffreys

(Received 1947 February 15)

Summary

If Saturn's rings were several particles thick the damping effect of collisions on the inclinations and eccentricities would reduce the rings in less than a year to a state where the particles were piled one on another. Dissipation by friction and impact would not cease at this stage. Its later effects would extend the ring inwards and outwards in its own plane, until it was nowhere more than one particle thick and the spacing was just enough for collisions to be avoided. The time needed to attain this state is estimated to be of the order of 10^6 years for particles of diameter 1 cm. and less for larger ones. Some comments are made on Maxwell's and Goldsbrough's criteria for stability.

The outstanding observational facts about Saturn's rings are as follows. (1) Spectroscopic evidence shows that the velocity at any point is nearly that of a particle in a circular orbit at the same distance. (2) The reflecting power is high, that of Ring B exceeding that of the planet and that of Ring A approaching it. (3) Nevertheless the rings are not quite opaque, stars having occasionally been seen through them. (4) The reflecting power is appreciably lower when Saturn is at quadrature than at opposition, even when allowance is made for the difference of distance. (5) The ring is extremely thin.

Maxwell showed that a set of satellites moving in one circle about the planet would be stable, and that all the other suggested types of constitution that he considered would be unstable. Since his essay his results have usually been quoted as the chief evidence for the meteoric constitution, but they are not quite decisive, because in considering a liquid or gaseous ring he supposed it to be in uniform rotation like a rigid body. Previous work of Laplace and Kowalewsky had shown that a ring in such a state of motion would have a thickness comparable with its width, and in view of our present knowledge of the thickness of the ring this hypothesis scarcely merits further consideration. But a fluid ring could be arbitrarily thin if we abandon the hypothesis of rigid-body rotation, which in itself would suggest a very high viscosity.* The stability of a fluid ring with variable rotation has not, I think, been discussed. The best way of presenting the case for a meteoric constitution at present, I think, is to rely on the observational data (1), (4) and (5) above. (1) and (5) limit us to the meteoric theory or to a gas or liquid with the velocities mainly controlled by gravity and not viscosity. A gas can be excluded, since the distribution of density normal to the plane would satisfy the usual laws for a gas. But in the small normal field that must exist in a mass 13 km. thick at most the density could not build up sufficiently to give great scattering of light. A liquid, again, would have a smooth surface and give regular reflection. Images of the ball of Saturn and of stars would be formed in the ring, and could not have escaped observation. Accordingly, quite apart from the mechanical arguments, which are incomplete, I think that optical considerations alone are

* For an analogous problem *cf.* Jeffreys, *The Earth*, 1929, pp. 49–52, Cambridge.

enough to show that the meteoric theory is the only tenable one. Further, the meteoric theory explains (4), as has been shown by H. Seeliger.*

Nevertheless the meteoric theory needs considerable mechanical investigation before it can be regarded as complete. Maxwell's essay was confined to a set of particles in a single circle, and recent investigations of perturbations have considered only disturbances from such a circle. Attractions between particles at different distances are ignored. But in any case, since the masses of the satellites are small, and that of the ring also small, all mutual gravitation is of the second order of small quantities. But in a system of freely moving solid bodies we may expect collisions to be frequent and to give discontinuous changes of momentum of the first order of small quantities. A preliminary treatment of these has been given †, but further investigation is needed.

The essential point is that the high opacity of the rings shows that on an average most rays of light striking the ring at angles up to $27°$ meet a particle on their way. But every particle of the ring must cross the mean plane of the ring twice in each revolution. Consequently we may expect it to undergo at least two collisions on the way. The point is that the opacity and the frequency of collisions depend on the same function of the number and size of the particles, namely the total surface of the particles per unit area in the plane of the ring; and if the departures of the particles from steady motion in circles can be treated as random the opacity shows that collisions will have a dominating effect. But collisions between solids are essentially non-conservative; at each collision the relative velocity is reduced by imperfection of restitution and by friction, usually by a fraction in the neighbourhood of $\frac{1}{2}$. Hence relative velocities between neighbouring particles will be rapidly annulled, with a time of relaxation not more than the orbital period, about a day. If for instance the particles were originally in an anchor ring, an average normal to the plane of revolution intersecting several particles, the velocities normal to this plane would be practically annihilated in a year. Further, the radial velocities corresponding to the orbital eccentricities would be removed at the same rate. But there is little to alter the mean motions. Hence the state reached in a year would be a peculiar one. The ring would be thin, but at any distance several particles would be piled one on another, in permanent contact. Such a state could last with little change for a long time, because the particles would acquire such rotations that there would be little difference of velocity at the points of contact. Nevertheless dissipation would not be altogether abolished. Detailed treatment is beyond the present resources of statistical mechanics, for even in the absence of dissipation the problem would be that of a fluid with the molecular spacing comparable with the molecular dimensions. But motion in and out, or up and down, would persist, though its actual amount would be only that needed to maintain rolling. Interchange between motions in different directions remains possible, since the lines joining centres of bodies in contact might be in any direction in relation to the mean plane of revolution. We have in fact a case of kinetic theory where there is no independent agitation; such agitation as there is would be parasitic on the general motion.

The outstanding cause of further dissipation would be the differences in orbital period. For simplicity take the particles to be spheres of radius a, and

* *Abh. Bayer. Akad.*, **18**, 1–72, 1893. For full discussion *cf.* E. Schönberg, *Handb. Astrophys.*, **2**, part I, 130–170, 1929.

† Jeffreys, *M.N.*, **77**, 89–92, 1916.

consider two sets moving in circles of radii r, $r+b$, where $b<2a$. In free motion the ratio of the periods is $1-\frac{3}{2}b/r$. Hence if the spacing in longitude is also b, a particle on the inner circle will encounter on an average $\frac{3}{2}$ particles in a revolution and share momentum with them. If v is the orbital velocity at distance r, m the mass of a particle, the outward transfer of angular momentum at a collision is of order $-mbr\,dv/dr$, the numerical factor being less than 1 but probably more than 0·1. Hence a particle transfers outwards, per unit time, an angular momentum of order

$$-\frac{3mbv}{4\pi}\frac{dv}{dr}.$$

As the particle occupies an area $2ab$, this is equivalent to a tangential stress

$$-\lambda\frac{mv}{ar}\frac{dv}{dr},$$

where λ will probably be between 0·01 and 0·1. Using the relation

$$v^2=\mu/r$$

we have for the tangential stress

$$-\lambda m\mu/2ar^3,$$

and for the total rate of outward transfer of angular momentum in a belt where the normal section is c,

$$\pi\lambda\mu cm/ar^2.$$

The total angular momentum in a ring, if a, b, c are constant, is

$$\tfrac{1}{3}\mu^{\frac{1}{2}}\frac{cm}{ab}[r^{\frac{3}{2}}].$$

Comparing these results, we see that the time that would be needed to transfer the whole of the angular momentum from the inner to the outer half of the mass would be of order

$$\frac{2}{3\pi\lambda}\frac{r^{\frac{1}{2}}[r^{\frac{1}{2}}]}{nb},$$

where n is the mean motion of a particle at the mean distance of the rings. If we take them as having their present extent, and $b=1$ cm., this quantity is of the order of 10^6 years. In the early state that we are considering, when the values of r would differ by much less, the time would be shorter. It would also be shortened by taking b larger, as we should have to do if we are to maintain the hypothesis that the particles were produced by tidal disruption of a solid body. The conclusion to draw is therefore that though the evolution in this state would be slow compared with the damping out of inclinations and eccentricities, it would still be rapid on a cosmogonical scale.

The nature of the changes is clear; on account of the steady outward transfer of angular momentum the outer parts of the ring would be driven outwards and the inner parts inwards, so that the ring would become thinner. The process would stop when the relation $b<2a$ is no longer satisfied. That is, we start from the state of close packing in three dimensions, and arrive at one where the particles are just widely enough spaced to avoid collisions altogether. They follow each other around in circles, the circles being spaced at intervals slightly greater than the diameters of the particles; and the ring is nowhere more than one particle thick.

The result is, I think, consistent with Seeliger's conclusion that a distance between particles decidedly more than their diameters is needed to explain the reduction of albedo near quadrature. It cannot be many times the diameters, since the high albedo even when the rings are open to their fullest extent shows that a straight line from the Earth to the ring must usually intersect at least one particle. If we were in a position to see the ring normally a much smaller fraction of the area would appear occupied than from our actual viewpoint.

Some suggestions arise from these results that may be relevant to the question of the sizes of the particles and to the stability. H. Struve * has inferred from the failure to detect secular perturbations due to the rings that the total mass is not more than 1/27000 of that of Saturn, say 2×10^{25} g. Supposing the whole area of the rings covered to a thickness $2a$ by matter of density 1, this gives $a < 1{\cdot}8 \times 10^4$ cm. This is less than we have inferred from the hypothesis of tidal disruption of a solid, and of course also less than the maximum thickness inferred from observation, which is of the order of 10 km.; and we seem to be driven to the additional hypothesis that the bodies produced in this way were broken up further by collisions in the early stage or to think of some different explanation altogether, such as that the rings were formed by direct condensation from the gaseous state as in the formation of snowflakes. Allowance for the fact that the whole surface need not be covered would not bridge the gap. The collisions in the later stage would be quite gentle, since the relative velocity of neighbouring particles would be of the order of their diameter in a day—much less than the velocity of fall of snowflakes.

According to Maxwell's theory, a set of particles moving in a circle would be stable only if $mp^3 < 2{\cdot}3M$, where p is the number of particles and M is the mass of Saturn, $5{\cdot}6 \times 10^{29}$ g. With spacing equal to the diameters we have

$$\frac{mp^3}{M} = \left(2\pi \,.\, \frac{1{\cdot}4 \times 10^{10}}{2a}\right)^3 \frac{4}{3}\pi a^3 / 5{\cdot}6 \times 10^{29}$$

which is very roughly 6. The result is independent of a; but if we allow for the spacing being greater than $2a$ Maxwell's criterion will be satisfied by the particles in each ring separately. It would not, however, have been satisfied in the earlier stages. But the stability in any case needs re-examination, on account of the effects of the attractions between particles in different circles.

If there was a division in the ring while the particles were still in contact, the tendency of the transfer of angular momentum would always be to fill it up. Consequently it is likely that the divisions in the ring have been formed since collisions became rare.

Goldsbrough has recently published † a detailed discussion of the perturbations of a circle of satellites by an independent satellite, and concludes that for certain ranges of distance the perturbations would produce instability; he finds that these ranges, for perturbations by Mimas, show interesting correspondences with the boundaries of the rings, Cassini's division, and Encke's division. The theory is very intricate and it seems hypercritical to suggest that it is not intricate enough, since it neglects the attractions between particles at different distances and the reaction on Mimas, whose mass is at any rate much less than the maximum possible mass of the ring. Goldsbrough works in terms of a parameter νT_s, which also

* H. Struve, *Publ. Obs. cent. Nicolas*, **11**, 228, 1898.
† *Phil. Trans.* A, **239**, 183–216, 1946.

appears in Maxwell's stability condition. Cassini's division, in particular, corresponds to the inequalities

$$2 \leqslant \Omega \leqslant \frac{2}{1 - \frac{7}{4}\nu T_s}$$

and ω/ω', the ratio of the mean motions of the particles and Mimas, is given by

$$\frac{\omega}{\omega'} = \frac{\Omega}{\Omega - 1}.$$

From the width of the division he derives a value of νT_s, which is consistent with Maxwell's criterion. His argument is then that a circle of particles not satisfying this condition and moving in the division would be perturbed until some of its members collided with particles in the main rings; the circle would then be broken up. But it seems to me that this only shows that for each departure of Ω from 2 there is a critical value of νT_s, below which the circle would be safe. The division would not be sharp; the brightness would fall off continuously towards $\Omega = 2$, but would vanish only at this value.

On Goldsbrough's theory the inner edge of Ring B and the outer edge of Ring A correspond to instabilities at $\Omega = \frac{3}{2}$ and $\Omega = 3$. It is natural to take the former as an indication that particles near the inner edge of Ring B would overshoot the danger zone and join the crape ring; but if so there will be a systematic loss of angular momentum which, as far as I can see, would be compensated by a gain by Mimas, which would thus be driven further off. But if we accept this explanation for the crape ring we should expect another crape ring outside Ring A. This is not definitely confirmed, though some observers have suggested it.

St. John's College,
Cambridge :
1947 *February* 14.

ON THE THEORY OF GLOBULAR STAR CLUSTERS

E. Finlay-Freundlich

(Received 1947 March 14)

Summary

The theorem is proved that a spherical system of stars contracting under the influence of its own gravitational field passes at one and only one moment, before reaching the state of greatest condensation, through a quasi-stationary state in which its structure corresponds to a solution of the equations of hydrostatic equilibrium.

The theorem is then applied to observational data discussed in a preceding paper.

1. *Introduction.*—In a preceding paper* I derived from the latest densitometric observations, which had revealed wide envelopes of faint stars surrounding the clusters of bright stars, that the structure of the globular star clusters corresponds very nearly to that of an isothermal core, containing the most massive stars, surrounded by an "adiabatic" atmosphere of light stars. Obviously a cluster cannot remain in this state for an unlimited time, because in the end isothermy should prevail and the cluster be scattered over an infinite space. The present structure therefore represents only an intermediate phase. The following paper adds considerable support to the theory of the preceding paper by showing that every cluster necessarily passes, in the course of its evolution, through such an intermediate phase. It will be proved that a spherical cluster of stars, contracting under its own gravitational field from a very diluted initial state, must pass, *before* reaching the state of greatest condensation, through a distinguished state, called here *quasi-stationary*, not because in this state the contraction comes to a standstill but because in this phase the internal structure of the cluster remains practically unchanged for a long time and resembles during this phase the structure to be expected, if the hydrostatic equations of equilibrium were directly applicable.

The following theory applies therefore only to the contracting phase in the evolution of a cluster which probably is only a relatively short phase of its total lifetime, though not necessarily short in comparison with the present age of the universe. It is an essential feature of the theory that purely *dynamical* considerations are used to explain the present structure of the globular star clusters, despite its close resemblance to the structure of a gas-sphere. This is essential, because it has been generally agreed † that the principles of the kinetic theory of gases are not applicable to globular star clusters. Owing to this, their actually observed structure remained hitherto unexplainable.

The following dynamical considerations show that the contracting phase of a spherical cluster of stars leads to a state in which its structure corresponds to a solution of the differential equations of hydrostatic equilibrium, without defining the actual structure more closely. The observations indicate that the solution attained is that of an isothermal core surrounded by an adiabatic atmosphere.

* *M.N.*, **105**, 237, 1945.

† ten Bruggencate, *Sternhaufen*, p. 97, Berlin, 1927.

After having reached this quasi-stationary state the cluster continues to contract, with only slight changes of its internal structure, until, owing to the increasing density of the core, encounters of stars become more and more frequent, and statistical considerations are needed to explain the further development which ends in isothermal scattering.

It appears to me of particular significance that a purely dynamical theory can be given which links organically the initial dynamical phase of contraction to the final statistical phase. The quasi-stationary state, considered here, is, so to speak, the branch-point connecting the two phases.

2. The following considerations start from a schematic model of a globular star cluster, supposed to be a finite spherical agglomeration of stars. In the differential equation from which the Virial Theorem is derived,

$$\tfrac{1}{4}d^2I/dt^2 = T + \tfrac{1}{2}V, \tag{1}$$

the cluster is characterized at every moment, t, by the value I of its moment of inertia:

$$I = \int_0^R \frac{dM(r)}{dr} r^2 dr,$$

where $dM(r) = 4\pi\rho(r)r^2dv$, $\rho(r)$ the density at the distance r from the centre, R the finite surface radius. T is the kinetic energy, V the potential energy of the cluster for a Newtonian gravitational field. We make the assumption that the cluster has been slowly contracting under its own gravitational field, all the time remaining spherical. Stars which have been lost by "evaporation" during this contraction shall be disregarded; thus the total mass of the cluster M remains constant. The cluster is said to expand, if $dI/dt > 0$ and to contract when $dI/dt < 0$; expansion being accompanied by a decrease, contraction by an increase in $|V|$.

Equation (1) is the "equation of motion" of the dynamical system, i.e. the globular cluster; it has the energy integral

$$T + V = E, \tag{2}$$

in which the energy constant E is supposed to be negative. With the help of (2) the kinetic energy T may be eliminated from (1), yielding the equation

$$d^2I/dt^2 = -2V + 4E. \tag{3}$$

Since V is an unknown function of r, not many results have hitherto been drawn from equation (3), apart from the Virial Theorem. It will be shown, however, that by giving to (3) a new formulation and by performing one integration, important conclusions can be drawn which apply to the evolution of globular star clusters.

3. The following special assumptions have to be made:

(*a*) During the contraction the value of the potential energy

$$V = -\int_0^R \kappa r^{-1} M(r) dM$$

shall be a *monotonic decreasing* function of the time t; for $t = 0$, i.e. at the initial stage of the evolution, I is supposed to have been very large and T small;

(*b*) The density

$$\rho = \frac{dM(r)}{4\pi r^2 dr}$$

shall be a regular function of r only; likewise the mean square velocity $\overline{c^2}$ of the

stars' velocities with regard to their common centre of gravity. Then we may define also the quantity $p(r) = \frac{1}{3}\rho \cdot \overline{c^2}$ as a regular function of r only; p measures in hydrostatics the pressure, but need be treated here as no more than a convenient parameter. No assumption is made concerning a special velocity distribution.

(*c*) There is no reason to doubt that the two preceding assumptions correspond to the physical conditions obtaining in a globular cluster. The special, but none the less extremely general, assumption that the expression

$$f(r) = -\frac{1}{\rho}\frac{dp}{dr} - \frac{\kappa M(r)}{r^2}$$

does not change its sign within the range of all values of $0 \leqslant r \leqslant R$ is also introduced for mathematical convenience. This assumption represents perhaps only a first approximation to physical conditions but merely implies that during the contraction the cluster is at every moment contracting at every point of its whole volume.

4. Equation (1) can be brought into a new form by introducing first into the expression of the kinetic energy

$$T = \int_0^R \tfrac{1}{2}\rho\overline{c^2}\, dv$$

the expressions

$$p = \tfrac{1}{3}\rho\, c^2 \quad \text{and} \quad dv = 4\pi r^2\, dr.$$

This gives

$$T = 6\pi \int_0^R pr^2\, dr,$$

which integrated by parts yields

$$T = 6\pi[\tfrac{1}{3} pr^3]_0^R - 2\pi \int_0^R r^3\, dp.$$

The first term vanishes, because $r = 0$ at the lower limit and $p = 0$ at the upper limit $r = R$; hence,

$$T = -2\pi \int_0^R r^3\, dp.$$

The potential energy, on the other hand,

$$V = -\int_0^R \kappa r^{-1} M(r) dM$$

can be written

$$V = -4\pi \int_0^R \kappa \frac{M(r)}{r^2} \rho r^3\, dr.$$

Introducing these new expressions for T and V into equation (1) produces

$$\tfrac{1}{4}\frac{d^2 I}{dt^2} = -2\pi \int_0^R \left[\frac{1}{\rho}\frac{dp}{dr} + \kappa\frac{M(r)}{r^2}\right] r^3 \rho\, dr. \tag{1.1}$$

5. In the preceding paper we put the bracket expression on the right-hand side of (1.1) equal to zero, thus obtaining an equation mathematically identical with the equation of hydrostatic equilibrium. By borrowing Emden's mathematical solution of this equation, we thus obtained a structure for the globular clusters

consistent with observations. We now show by using the equation of motion of the cluster in the forms (1.1) and (3) that during its evolution each cluster *must* pass through a state in which d^2I/dt^2 vanishes. In such a case the bracket-expression in (1.1) vanishes and the cluster's structure satisfies the equation of hydrostatic equilibrium

$$\frac{1}{\rho}\frac{dp}{dr} = -\frac{\kappa M(r)}{r^2}.$$

More precisely, it will be proved that, preceding the moment of greatest condensation, there is one and only one instant $t = t_1$ at which $d^2I/dt^2 = 0$. In order to prove this theorem we start from equation (3):

$$d^2I/dt^2 = -2V + 4E.$$

The first term on the right-hand side is positive; the second term is a negative constant.

After performing one integration we obtain

$$\frac{dI}{dt} = -2\int_0^t V(\tau)d\tau + 4Et + c,$$

where c is a new constant of integration.

We shall first consider the case $c = 0$; the cluster is then supposed to have been contracting from the beginning solely under its own gravitational field. For sufficiently small values of t, dI/dt and d^2I/dt^2 will both be negative and the cluster will contract.

But since $-V$ is a monotonic increasing function of the time t, the mean value

$$-\bar{V} = -\frac{1}{t}\int_0^t V d\tau$$

is also a monotonic increasing function of the time and thus there must exist a value $t = t_0$ for which

$$-2\int_0^{t_0} V d\tau = -2\bar{V}_{t_0} t_0 = -4Et_0$$

and for which therefore

$$dI/dt = 0.$$

Beyond this moment t_0 we do not investigate the contraction of the cluster.

Again, in accordance with the "mean value theorem for integrals" there must exist a value t_1, $0 < t_1 < t_0$ for which

$$-V(t_1)t_0 = -\int_0^{t_0} V\, d\tau,$$

so that

$$\frac{dI}{dt} = -2V(t_1)t_0 + 4Et_0 = 0$$

or

$$t_0[-2V(t_1) + 4E] = 0.$$

Consequently for $t = t_1$,

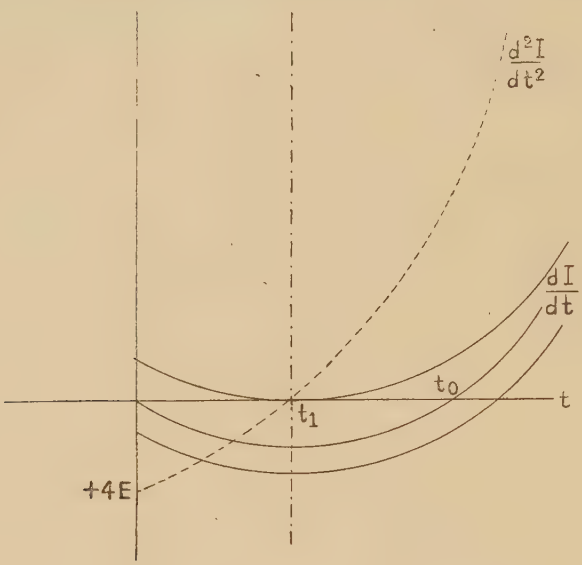
$$d^2I/dt^2 = -2V(t_1) + 4E = 0,$$

and therefore also in accordance with (1.1)

$$\frac{1}{\rho}\frac{dp}{dr} + \kappa\frac{M(r)}{r^2} = 0.$$

At this epoch, preceding the moment of greatest condensation, the cluster passes through a quasi-stationary state.

6. The general case $c \neq 0$ is best treated in a geometrical way. In the diagram the abscissae measure the time t; the dotted curve illustrates in a schematized way the values of d^2I/dt^2 as a function of the time; the other curves give the values of dI/dt, each curve corresponding to a different value for the constant of integration c. Since d^2I/dt^2 is a monotonic increasing function and starts for $t=0$ from the negative value $4E$, the curve for d^2I/dt^2 must pass through the t-axis. This occurs, as has been proved in the preceding paragraph, for $t=t_1$. The dI/dt curve, for $c=0$, has on the other hand for $t=t_1$ a horizontal tangent and passes the t-axis at a point $t_0>t_1$. Similarly all dI/dt curves have, when $c\neq 0$ at $t=t_1$, a horizontal tangent; dI/dt reaches here its lowest value and from there on the value rises again.



For negative values of c (only such values are compatible with contraction in the initial phase) the curves pass in one and only one point $t=t_0>t_1$ through the t-axis. When $c>0$, but sufficiently small, the curves pass twice through the t-axis, once for $t<t_1$ and once for $t>t_1$. This is the case for increasing values of c until for $c=c_0$, where

$$c_0=\int_0^{t_1}[2V(\tau)-4E]d\tau,$$

the t-axis itself becomes at $t=t_1$ a tangent to the dI/dt curve. We consider here only the cases $c\neq 0$; in all such cases there is one and only one instant, preceding the epoch of greatest condensation, at which the cluster passes through a quasi-stationary state, i. e. a state in which its density distribution corresponds to a state of hydrostatic equilibrium.

7. Although the theorem just proved refers to a simplified schematic model of a globular star cluster, it is general enough to be applied to the evolution of actual globular clusters.

The observations indicate that the globular clusters are in or near a quasi-stationary state consisting of an isothermal core surrounded by an adiabatic atmosphere. We may therefore conclude that they are approaching or have

actually reached the phase when d^2I/dt^2 vanishes or at least is very small. The contraction of the cluster will probably proceed at a "cosmic" time-scale, which implies that dI/dt will be all the time very small. Thus, near to the state of greatest condensation, when d^2I/dt^2 becomes zero, dI/dt and d^2I/dt^2 will both be small, and the cluster will undergo during this whole period only very small changes; its structure therefore will correspond for a relatively long time to a solution of the equation for hydrostatic equilibrium. This explains probably the fact that we find the globular star clusters resembling each other to a high degree. If in addition they differ from each other only slightly in total mass and energy, it becomes intelligible that the absolute values of their radii R scatter only very moderately around a mean value as the observations show.

An important problem arises from the question whether the isothermal-adiabatic state is the one distinguished state through which every cluster must pass before the final disintegration begins. The answer to this question can only be given when we have obtained a full knowledge of the dynamics of the contracting phase. However, since isothermy must be the final state, the development of an isothermal core which, during one special phase of the evolution, is just able to "support" the overlying envelope of stars may be a significant intermediate phase that always has to be passed.

I am very much obliged to Dr Walter Ledermann for his valuable advice which I followed for instance in the geometrical considerations in Section 6. But I am particularly indebted to Dr G. Leslie Camm who went through the whole theory and discussed critically its results with me before the final manuscript was drafted; Mr Ian Campbell gave me valuable help in formulating the paper.

University Observatory,
St. Andrews :
1947 *March* 12.

THE FRAUNHOFER LINES OF THE SOLAR SPECTRUM

(George Darwin Lecture, delivered by Professor M. G. J. Minnaert, on 1947 May 9)

A lecture on Fraunhofer lines has inevitably a somewhat abstruse character. The spectroscopist could not show you any of the wonderful pictures which the telescope reveals. A spectrum looks at first sight like a rather dull succession of bright and dark stripes. But now comes the observer, using his refined methods and discovering the infinite variety of shades and halftones in that spectrum, the full richness of that chiaroscuro. And then comes the theorist, deriving by the power of his phantasy how that music of undulating spectral lines is due to the airy dance of the atoms in the fiery radiation of the Sun. Actually, the investigation of Fraunhofer lines is so fascinating, so rich in observational details and so important theoretically, that I shall have to restrict myself to one part of the subject: photometry; even then, I shall only be able to present before you some of the more important moments in the development of our knowledge and a general view of the present state of the problems involved. In this frame I shall give an account more particularly of the work done at Utrecht because with this work are connected so many personal reminiscences; but we all know that progress in this field is due to the cooperation of many scientists all over the world.*

Already the first observations of a good solar spectrum revealed, that among the individual Fraunhofer lines there is a great variety of intensity profiles. When Kirchhoff in 1860 made his drawing and catalogue of the solar spectrum †, a work in which he lost most of his sight, he with great care gave to each line a characteristic letter from *a* to *g*, indicating the *width* in 7 different grades; and similarly he had also a scale for the *darkness* inside the line, running from 1 to 6. His lithographic reproduction of the solar spectrum was printed with 6 different stones, each of them corresponding to the lines of one definite degree of darkness. It is interesting to find the green magnesium line $\lambda 5183$ as a broad, shallow band with a sharp central core.

Rowland, in his Preliminary Table of Solar Spectrum Wave-lengths (1893–1896), characterized each Fraunhofer line by a number, the "intensity": a preliminary, rather qualitative estimate of the impression which the line makes, either by its width or by its darkness or by both of them. This celebrated Rowland scale, though being admittedly rough, proved of the greatest use during many years up to our day. Rowland's collaborator, Jewell, made interesting observations on some profiles.‡

Actual measurements concerning the intensity inside Fraunhofer lines were not made before 1900. An enormous amount of material was available, a vast field would be open to investigation, as soon as quantitative measurements could be made. The pioneer in this field, as in so many others, was Schwarzschild at Göttingen. His profiles of the lines H and K, determined for the centre of

* General references to the literature may be found in Unsold, *Physik der Sternatmosphären*, and in the report by Redman, *M.N.*, **98**, 311, 1938.

† "Untersuchungen über das Sonnenspectrum", *Abh. Akad. Berlin*, 1861.

‡ *Ap. J.*, **3**, 100, 1893.

the Sun's disk and for the limb, are the first profiles ever measured (1914); their precision is still comparable to that of modern measurements.

But Schwarzschild's photometric method is cumbersome, and for many years nobody followed the new pathway, discovered by him. It was only in 1927 that the tradition of reliable photographic photometry, which had been kept at Potsdam, was applied to the investigation of Fraunhofer lines by the younger generation of astronomers and physicists.*

The independent development of the Utrecht measurements† was due to the happy circumstance that Julius, professor of experimental physics, had created a heliophysical department at his institute. Since 1920 our physicists, under the lead of Ornstein, directed their efforts to the photometry of laboratory spectra; Moll constructed his microphotometer; Burger and Dorgelo discovered the multiplet rules; the photometric properties of photographic plates were investigated; van Cittert and Burger used the step reducer and developed a method for the correction of a profile for instrumental influences. It was very stimulating for an astrophysicist to take part in that development, and to apply the wonderful new methods to the investigation of Fraunhofer lines. The heliophysical department found at once a new field of activity. And I remember that our very first microphotometer record of some small part of the solar spectrum awoke at once the very ambitious plan of a complete photometric atlas. Our first steps were very cautious. In a spectrum, obtained with a focal length of many metres, we had to record the detailed profile of dark lines, not wider than a few tenths of a millimetre, surrounded by floods of light; and we wanted to measure the intensity in each point with a precision of a few per cent! It seemed almost incredible that the intensity profile recorded would have any real value. All possible sources of error were examined one by one, their influence on the measurements were minimized and the remaining effects were estimated. Gradually, we found that our results were more reliable than we could have hoped and our self-confidence increased. It was a great encouragement, that soon similar measurements were made at other institutes all over the world, and that the results from this pioneering period seemed to agree in a rather unexpected and very satisfactory way. When reading again the papers from this first period by von Klüber, Unsöld, Plaskett, Dunham, Righini, Woolley, Carroll and those of the next generation, represented by Redman, ten Bruggencate, Thackeray, Allen, and at Utrecht by Mulders and Houtgast, one is greatly impressed by their fundamental work, which became the basis for all modern spectrophotometric investigations. It may be safely asserted that astrophysicists have followed the good tradition of classical astronomy, and that they have bestowed all possible care on their method before they started measuring.

Our reward is, that we now see our naïve younger students making without any hesitation these same measurements, which for us were such a source of trouble and concern, and getting satisfactory results!

The number of Fraunhofer lines of which the profile has been studied with some precision is still astonishingly small. I mention the first four Balmer lines, H and K, the *Ca* resonance-line 4227, the D lines of sodium, the *Sr*-ion line 4078, and five heavy lines of iron in the violet, altogether sixteen lines. You will understand, that we felt a very strong desire to make available for theoretical investigation

* von Klüber, *Z. Phys.*, **44**, 481, 1927.
† *Z. Phys.*, **45**, 610, 1927.

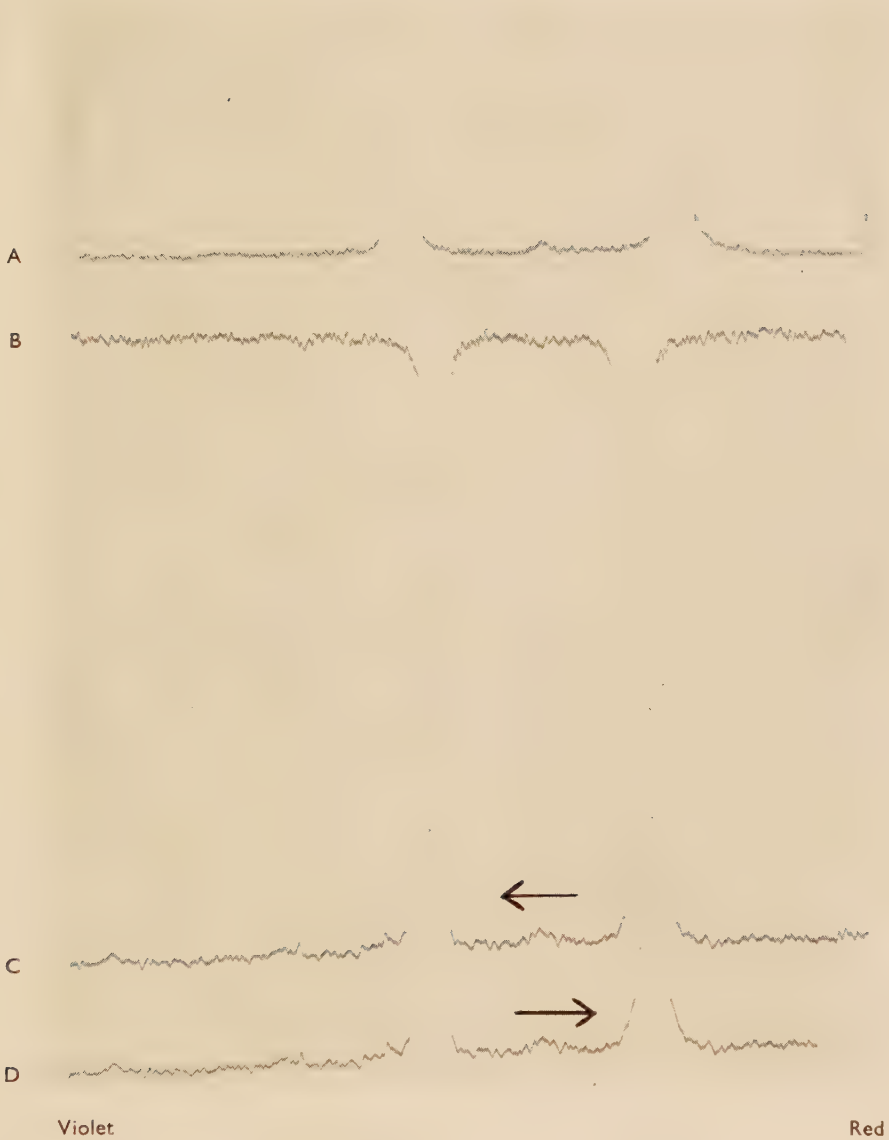
not only these sixteen lines out of 20,000, but many more, and even all of them, if possible. The plan for a general photometric atlas of the whole solar spectrum gradually took shape and seemed more and more possible through the experience, accumulated at Utrecht and increased by information from other workers in the field. Of course, in such an undertaking it is impossible to give full care to each of the lines, as may be done in individual investigations. But if the work is done in a systematic way, one may be able to correct afterwards in a relatively simple way the errors left. The plan was carried out by Houtgast, Mulders and myself, each taking for his share a part of the work.* A great deal of the credit is due to the generous help of the Mt. Wilson Observatory, where Mulders was allowed to take the whole series of spectral records with the splendid equipment there available. The spectra have a dispersion of 3 mm./A. in the second order, beautiful definition, and a reliable standardization; their full length is about 12m. In the very first period of spectrophotometric research, one would have had to measure by eye the density of all consecutive surface elements of the plate by means of a Hartmann photometer. Since the invention of the microphotometer, this whole process of deriving the transmission of the plate all along the spectrum is solved by an entirely automatic apparatus. However, the second step involved, the conversion of the transmission curve into an intensity curve, is still now generally made by tedious measuring and reading from graphs and plotting point by point, exactly in the same cumbersome way as we formerly did when we determined transmission with the Hartmann. In our case, it would have been utterly impossible to follow such a lengthy procedure. The total length of the microphotometer records amounted to 120m. When handling such extensive material, a completely automatic and direct registration of the intensity curve is necessary.† This was solved by adding a relatively simple auxiliary apparatus to the ordinary microphotometer; a diaphragm is inserted, cut out in the form of a characteristic photographic curve, corresponding to the plate which we are investigating. The instrument works remarkably well and has been in full operation for the whole of our atlas. For a similar purpose an entirely different instrument of great ingenuity has been devised by Williams and Hiltner at Michigan; it was used by them for their wonderful Atlas of Stellar Spectra. I have the impression that both of these instruments are adapted to their particular problem. The Michigan apparatus can be used without complications even if the characteristic curve changes quickly along the spectrum, as is generally the case in stellar spectra of moderate dispersion; but it has the disadvantage that it increases the graininess of the plate. The Utrecht device is only practical when the characteristic curve remains the same over a considerable distance along the plate, so that the same diaphragm can be kept in use, but it reduces the graininess to a minimum.

It was a real thrill to develop the intensity curves recorded, to see in the red light of the darkroom how the graphs appeared and how the profiles of the well-known Fraunhofer lines were revealed one by one for the first time in their true intensity values.

When turning the leaves of the atlas, many striking features of spectral lines become at once apparent. The shallow hydrogen lines and the Ca^+-lines in

* Minnaert, Houtgast, Mulders, *Photometric Atlas of the Solar Spectrum*, Amsterdam, 1940.

† Minnaert and Houtgast, *Z. Astrophys.*, **15**, 354, 1938.



A—Mount Wilson. B, C, D—Evershed.

the infra-red, the broad wings of the sodium and magnesium lines, the deep iron lines in the violet, the complicated blends in the ultra-violet are striking. Shortly after the publication, Mulders described the faint ultra-violet hydrogen lines up to H17, which Unsöld also remarked; they can only be seen on the microphotometer records and not by looking at the original plate.*

It is very important to examine the narrower lines of the atlas and to find out how far their profiles have to be corrected for instrumental distortion and broadening. From plates taken for that purpose, we determined the instrumental curve: we found that a truly monochromatic line would be broadened into a line with a width of 61mA. (in the second order, $\lambda = 6000$ A.). However, van de Hulst, from a careful analysis of the atmospheric oxygen lines, showed that our instrumental width had been overestimated a little, and that the width should rather be taken as 51mA.† Let us now consider the influence of this effect upon the narrow solar lines. From a random selection of well isolated lines, recorded in the atlas, we draw typical mean profiles. To correct these for instrumental broadening, several methods are available. In our experience, that of Burger and van Cittert ‡ is still very reliable; but van de Hulst § has described other and quicker procedures. We find now the corrected profiles as drawn in Fig. 1.

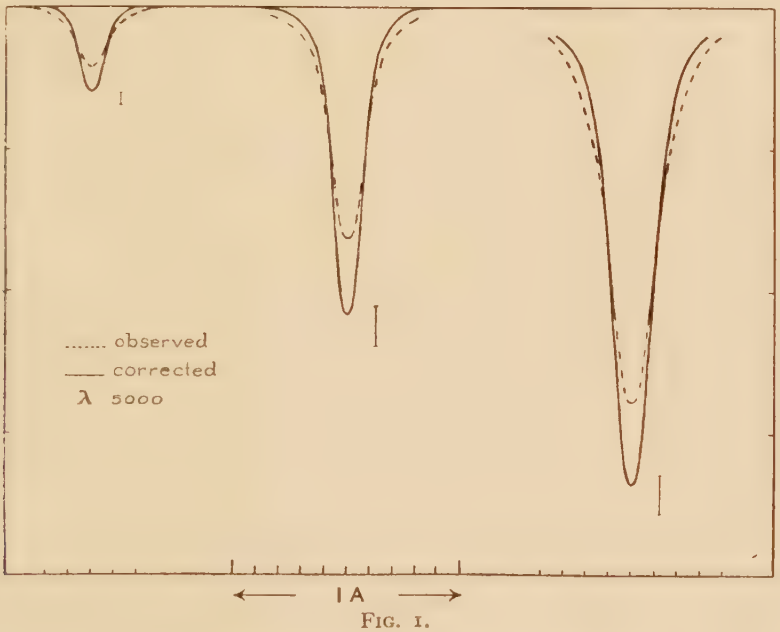


FIG. 1.

Profiles of faint Fraunhofer lines from the Utrecht Atlas.

A valuable check on these results is obtained from measurements by Shane ‖, who carefully determined with his interferometer the true central intensity of some weak lines. By comparison with the Utrecht atlas, he finds that the correction varies somewhat from plate to plate; according to his curves, the central

* Mulders, *P.A.S.P.*, **53**, 38, 1941.
† *B.A.N.*, **10**, 79, 1946.
‡ *Z. Phys.*, **79**, 722, 1932 and **81**, 428, 1933.
§ *B.A.N.*, **9**, 225, 1941; and **10**, 79, 1946. See also: Kremer, *B.A.N.*, **9**, 229, 1941.
‖ *P.A.S.P.*, **53**, 200, 1941.

intensities of the lines considered should fall within the range indicated in our diagrams by vertical bars (Fig. 1).

The excellent agreement shows that such corrections can be made in a quite reliable way. If the corrected central intensity is about right *, the other points of the profile will be still better. It is well known that in principle, given an accurate instrumental curve and precise measurements, arbitrarily narrow lines can be fully resolved. In practice there are very definite limits. But apparently the instrumental curve of our atlas is sufficiently narrow, compared to the rather considerable width of all solar lines, to make a reliable correction possible for all of them. The profiles thus obtained, similar to those drawn on Fig. 1, are the *true* shapes of the faint Fraunhofer lines. The atlas is sufficient for a full resolution of all solar lines.

The origin of the instrumental broadening is still not fully understood. For a part it is due of course to diffraction by the finite size of the grating; but other effects must be certainly involved also. As far as present experience goes, the width of this instrumental curve seems to be proportional to the wave-length, as might be expected for a diffraction phenomenon. It would be important to check this for different spectrographs.

Another interesting problem is the slight asymmetry toward the red, observed in a great number of the isolated lines throughout our atlas (Plate 9A). It is of course necessary to ascertain whether this effect must be ascribed to a real asymmetry in the Fraunhofer lines, or to an asymmetry in the spectrograph, or to inertia in the microphotometric apparatus.

1. We begin by recording a Mt. Wilson plate near $\lambda 6100$ with an ordinary microphotometer, without cylindrical lenses or intensity apparatus, just in order to have clean and simple experimental conditions. The asymmetry shows again, towards the *red*, apparently as strong as in the atlas.

2. The microphotometer is run in the reverse direction; the asymmetry remains, always towards the *red*; the records forward and backward are identical. This proves that inertia of the microphotometer is excluded (*cf.* Plate 9C and D).

3. We now investigate a plate of the same spectral region, taken with the grating spectrograph at Hale's private observatory, Pasadena. The asymmetry is again present, to the *red*, and by about the same amount as in the Mt. Wilson plate. One would almost suspect that the phenomenon is a true solar effect!

4. Fortunately, Evershed was so kind as to send us a few very fine plates, taken by means of his autocollimating spectrograph with liquid prisms, its dispersion is equivalent to that of ten 63 deg. flint glass prisms. Here the asymmetry in the spectral lines is small, but distinctly toward the *violet* side. (Plate 9B). In a second series of plates from Evershed the asymmetry is still more reduced (Plate 9C and D).

5. From these records it is now clear that the asymmetry in our atlas has an instrumental origin, which is to be found in the spectrograph. Let us now compare a spectrum taken at Hale's laboratory with the left side and another with the right side of the grating. The records show that the asymmetry is towards the *red* in both cases. It may be that the grating itself is the source of the trouble. It may be that in big spectrographs the objective does not always

* The central intensities found agree reasonably well with those of ten Bruggencate, *Z. Astrophys.*, **18**, 327, 1939. Those of Allen are appreciably lower, *Ap. J.*, **85**, 165, 1937.

concentrate the rays in a perfectly symmetrical image. In any case it is clear that all care should be given to such instrumental imperfections. A true asymmetry of Fraunhofer lines, mentioned sometimes in the literature, cannot be considered as established unless confirmed by very precise experiments.

It must be possible to account for the asymmetry of the atlas lines by considering the influence of the instrumental curve.* This curve, as it was determined from neon lines and published in the introduction of our atlas, shows indeed some asymmetry; a slight increase of this asymmetry would already be sufficient to explain the asymmetric line-wings; and this modification seems reasonable, considering the difficulties of determining an instrumental curve.

Fig. 2 finally presents the best available mean curve. It embodies: (*a*) our original determinations from neon lines; (*b*) a slight contraction after van de Hulst; (*c*) the asymmetry, as derived from the wings of well isolated lines.

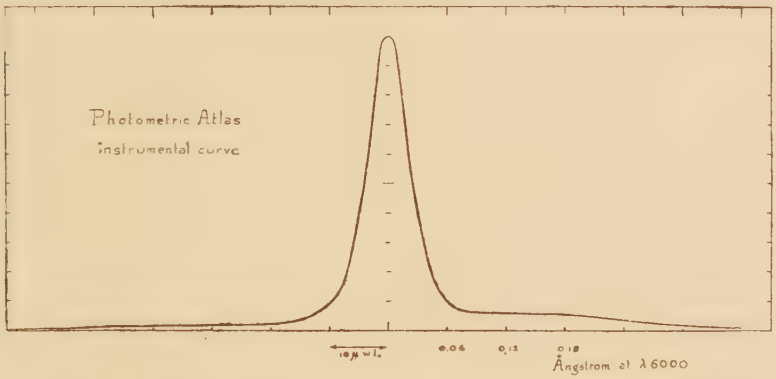


FIG. 2.

It will certainly prove very interesting to undertake a detailed investigation of solar line profiles either by means of the interferometer or by precise corrections to the atlas curves. For only this method gives information on the *individual* shape of each particular line, independent of any theory.

Apart from such a detailed and always difficult analysis, it is important that from each profile, however distorted by the instrument, we are able to derive a most valuable characteristic: the area, which is a measure for the energy absorbed in the line. This area, which we express as an equivalent width†, is a more precise definition of what was formerly called rather vaguely the "line intensity". The introduction of this concept proved very useful, both from a descriptive and a theoretical point of view. Already in 1934, Allen presented a beautiful catalogue of equivalent widths, which was extensively used.‡ After the publication of our atlas, we thought that it would be useful to derive a similar catalogue from our graphs. This work was undertaken in the difficult years of the war by a team of older and younger Utrecht astronomers, and I can inform

* This subject has been fully investigated by Messrs Buys and Ongering; I thank them for a preliminary communication of their results, which are being published in *B.A.N.*

† First introduced: *Versl. ned. natuur-geneesk. Congres*, 1923. From theoretical results of Allen and Menzel, it proved afterwards useful to measure widths in micro-wave-lengths $= 10^{-6}\lambda$. (Dunham, *Publ Amer. astr. Soc.*, **7**, 215, 1933.) This is the more so, because the width of the instrumental curve is also proportional to λ.

‡ *Mem. Commonw. Obs. Austr.*, **1**, No. 5, 1934 and **2**, No. 6, 1938; *Ap. J.*, **88**, 125, 1938.

you that the infra-red spectrum, recorded in the first order, is now practically ready for publication. This is the part most urgently needed, because it has not yet been treated by Allen and supplements his work. It is curious how many difficulties have to be solved in an apparently so simple undertaking. Thus the faint lines, with equivalent widths below 20mA., are difficult to measure with sufficient precision by means of a planimeter. We found a much better and quicker method in making glass scales with a great variety of standard profiles, once for all accurately measured, and to which the atlas lines are easily matched. Another very frequent difficulty is the question whether a small inflexion in the graph is due to a Fraunhofer line or to graininess of the photographic plate. Then for many strong lines there is doubt concerning the precise run of the continuous background. Abnormally broad lines may be considered as single and widened, or as a close doublet of normal lines. The analysis of blends is again a special problem; for this purpose, well isolated lines of the atlas were corrected, theoretically compounded into a great variety of blends, and finally blurred again by the instrumental curve into model blends, thus furnishing reliable standards for comparison. So there are many questions involved which can be solved only by much experience and discrimination.

As Miss Farnsworth at Mount Holyoke kindly offered the help of her observatory for this programme, we asked her to make independent measurements of several series of selected dubious lines. At the same time I am informed, by the courtesy of Babcock, that a very complete catalogue of the infra-red solar spectrum has nearly been finished at the Mount Wilson Observatory. Both sets of American measurements will furnish an excellent check on our catalogue and a very useful estimate of the possible errors, due to differences in interpretation.

In all these considerations we have mentioned "the" equivalent widths of Fraunhofer lines, as if their intensities were invariable and once for all fixed. It is absolutely necessary to investigate whether the solar spectrum is really constant, by comparing plates, taken with the same instrument at different times during a series of years. Such a fundamental investigation was made by van der Meer*, using the collection of spectrograms taken at Utrecht. His method, being purely differential, is at the same time very simple and very sensitive; it can be used for plates without any calibration. I would like to see it applied to the material of other observatories. The result was: an astonishing constancy of the solar spectrum, which means also constancy of the solar temperature; a difference of 25 deg. would have been detected, by comparing lines of very different excitation potentials. This is fully in line with the well-known constancy of the solar radiation. The variations in the far ultra-violet, ascertained from geophysical observations, apparently do not affect the gases which absorb the Fraunhofer lines, probably because the far ultra-violet radiation originates in the chromosphere. At the same time this is a very happy circumstance for all of us who are engaged in measuring intensities and making catalogues, because now we are sure that the equivalent widths have fixed values; values found at different times and by different observers can be directly compared, and the care given to precise measurements is justified.

While the methods of measurement were improved and while a huge amount of data accumulated, the theory also of Fraunhofer lines was developed, and it

* *Variaties van Fraunhofer-lijnen*, Dissertation, Utrecht, 1936; *Z. Astrophys.*, **13**, 157, 1937.

has reached at this moment a more or less classic form. For the sake of a systematic exposition, I have described first the observations, and I will consider now the theoretical side of the subject. Actually both methods of investigation have continually influenced each other and their interaction has been a great stimulus towards progress. Already the first and most primitive intensity measurements have contributed to the solution of that fundamental question: "how does a Fraunhofer line actually originate?" It is perhaps not generally known that Kirchhoff and Stokes each stressed part of the phenomenon. When Kirchhoff assumes that the solar gases have a greater absorption coefficient for light inside the line, and also a greater emission coefficient, he apparently means that there is purely thermal absorption: the absorbed radiation is converted into heat, and reversely the atom is again excited by collisions. But when Stokes pictures the vibrating atoms and resonators which re-emit the arriving waves *, he has in mind the process which we now call selective scattering: the absorbed radiation is at once re-emitted, without an interaction of other atoms. Kirchhoff's celebrated experiment on the reversal of the sodium lines shows the effect of selective absorption and emission, because it is made at a relatively high pressure and in the presence of many molecules. Stokes was not able to illustrate his view of the matter by an experiment; actually it was only in 1924 that Dunoyer by means of modern vacuum technique demonstrated the pure resonance radiation of sodium. It is interesting that these two theories even now each hold part of the field, and that their relative importance has not yet been entirely settled. For a long time the conception of Kirchhoff generally prevailed. The model of Stokes was taken up much later by Julius, Stewart and Schwarzschild and is now considered a better description of the phenomena.† That scattering plays an important rôle became at once clear when some measurements of profiles became available, made by Schwarzschild, von Klüber and at Utrecht. By thermal absorption the core of a Fraunhofer line could never become darker than 30 or 20 per cent of the continuous background, whereas many lines have central intensities of only 5 per cent and less.‡ Unsöld made a great step § by showing quantitatively that *scattering by damped atoms* accounts in broad lines for the profile of a strong Fraunhofer line. This profile is practically a resonance curve, one of these well-known curves which our students have been recording for many years in practical work on electrical circuits. These same curves now explain also why Fraunhofer lines are not sharp and black, but broad and hazy.

Originally only very few lines could be measured in a reliable way. It was important to make the bulk of all the fainter lines useful for theory, avoiding at the same time all difficulties and uncertainties of the instrumental curve and the correction for finite resolving power; this was the more necessary, because it was proved soon that the most interesting lines had only an equivalent width of a few hundredths of an angstrom. This problem was solved by making use of the equivalent widths, and when the relation between these equivalent widths,

* *Phil. Mag.*, **19**, 196, 1860.

† Selective scattering is identical with the anomalous scattering advocated by Julius. His work remains of historical importance, though the anomalous refraction in density gradients, also considered by him, is now known to be negligible.

‡ Thackeray, *M.N.*, **95**, 293, 1935; Redman, *M.N.*, **95**, 742, 1935 and **97**, 552, 1937; Allen, *Ap. J.*, **85**, 165, 1937; Shane, *Lick Obs. Bull.*, **19**, 119, 1941.

§ *Z. Phys.*, **44**, 793, 1927.

from the very faintest to the very strongest lines, was found. The method was introduced by Russell and proved wonderful.* Comparing the components of multiplets, we find a *curve of growth*, indicating how the lines increase in strength when the concentration of the resonators or of the absorbing atoms augments. Russell had considered the increase of the Rowland intensity. It was clear that more significant data could be obtained by plotting the equivalent widths.† We found remarkable curves, drawn at first in a purely empirical way, but which we at once recognized when, just in time, a paper of Schütz was published on the curves of growth of laboratory lines.‡ Curves of growth are very insensitive to the stellar model, but they give full information concerning the profile of the spectral lines. Just as these profiles can be deduced from a blurred spectrum by correcting for the instrumental curve, similarly they may be found from the curve of growth, in both cases by solving an integral equation. From our curves, Mulders and I were able to conclude that the damping of Fraunhofer lines is about ten times the classical damping, and that apart from this there is also a broadening due to the Doppler effect of thermal motion and appearing especially in the faint lines. The curve of growth for solar lines has been constructed later on the basis of much more material and with considerably increased accuracy by Righini, Menzel, Allen, R. B. King, K. O. Wright, and our original results have been mainly confirmed.§ From this work and from that of many other observers and theorists, it is possible now to picture already in detail the behaviour of the atoms in the radiation field of the Sun.

Let us consider a volume element of the solar atmosphere, containing some milliards of corpuscules. These we have never observed individually, we only hear the ether music emitted by the buzzing swarm, but the refined analysis of these waves reveals each motion, each adventure happening to any single atom. We see before our eyes that there are about 10,000 hydrogen atoms against one metal atom; more uncertain is the abundance of helium, carbon, nitrogen, oxygen, which together perhaps may be put equal to 2000 particles. The metals are for a great part ionized; a few per cent of them are excited. Some free electrons shoot through the swarm, about one to the thousand particles, most of them having been ejected from hydrogen atoms. All are moving quickly by thermal agitation. Moreover there are turbulent motions of the order of 1 km. a second. When ether waves arrive, these atoms start resonating; but why are they damped in doing so? This question has stimulated extremely interesting investigations on atomic behaviour. We first considered classical radiation damping: a string gives only a pure tone when vibrating in a vacuum, but then I cannot hear it; as soon as I perceive a sound, the chord must have transmitted energy to the air; so it has experienced some friction; so it is damped; so its tone is not pure any more. The curve of growth showed that the damping in Fraunhofer lines is actually ten times the classical radiation damping. Now an atom is not identical with a classical resonator. We calculated what

* Russell, Adams, Moore, *Ap. J.*, **68**, 1, 1928.

† Minnaert and van Assenbergh, *Z. Phys.*, **53**, 248, 1929; Minnaert and Mulders, *Z. Astrophys.*, **1**, 193, 1930 and **1**, 300, 1930; Minnaert and Slob, *Proc. Acad. Sci. Amst.*, **34**, 543, 1931. General review in *Observatory*, **57**, 328, 1934.

‡ Schütz, *Z. Phys.*, **64**, 662, 1930.

§ Righini, *Mem. Oss. Arcetri*, Fasc. 48, 1933; *Mem. Soc. astr. Ital.*, **7**, 1933; Menzel, Baker, Goldberg, *Ap. J.*, **87**, 81, 1938; R. B. King, *Ap. J.*, **87**, 40, 1938; Allen, *Mem. Commonw. Obs. Austr.*, **1**, No. 5, 1934; K. O. Wright, *Ap. J.*, **99**, 249, 1944; Rubenstein, *Ap. J.*, **92**, 114, 1940.

a quantum atom would do; it could be sometimes more, sometimes less damped than a resonator, but a factor 10 could certainly not be expected.

The other possibility was collisional damping. Physicists and astronomers are familiar with the idea, that the spectral lines become hazier by pressure. But could this effect operate at 0·01 atm.? And if so, how precisely does it work? * The kind of disturbance mostly considered by the physicists is a typical collision, which quenches the energy of the excited atom. But this is identical with what we called thermal absorption, and which we have already excluded because of the central intensities. Then Weisskopf and Lenz discovered another, a much more subtle disturbance.† Particles passing along had only to modify the frequency of a vibrating atom during a very short moment now and then, in order to confuse the phase coherence of the scattered wave-train. So this was the solution; there could be disturbance and still no quenching of the energy! The precise theory of this effect was developed later by Lindholm ‡ and in a more general form by van de Hulst.

But we ask further. *How* does a passing atom accelerate or retard the vibration? Are all atoms equally effective in this respect, and how effective are they? Here also, different possibilities have been considered. It may be that the quadratic Stark effect of the electrons plays a rôle. But Strömgren, calculated from atomic theory §, and ten Bruggencate and Houtgast || showed from observations that in the Sun the most important effect is due to the Van der Waals forces of the hydrogen atoms, forces which act proportionally to the inverse sixth power of the distance. This then seems to be the ultimate explanation for the haziness of Fraunhofer lines: *phase disturbances by hydrogen atoms.*

To the hydrogen lines themselves this theory does not apply. They are different from all other lines, because hydrogen atoms are the only ones which experience a linear Stark effect. In their case the statistical broadening by the Stark effect of neighbouring ions is the most efficient perturbation.

Finally there is the important fundamental question, whether after each absorption by an atom there follows an emission in the same frequency. A physicist would be astonished to hear the question put that way and he would reply: "Of course not; the returning electron can fall back just as well in steps as directly". Astrophysically speaking such processes are nearly irrelevant, because they are to a great extent balanced by the reverse processes; for nearly each quantum of light, split into pieces, there is a new quantum welded together. But owing to the anisotropy of the solar radiation field, the balance is not perfect, and a small amount of continuous radiation is pouring down into the strong Fraunhofer lines, so that their cores become less dark.¶

Also in another way there could be a difference between the radiation absorbed and the radiation emitted. We know that the atomic energy levels are not perfectly sharp, but may be pictured as groups of numerous narrow-spaced sub-levels. An electron, starting from one of the sub-levels, may fall back to

* A general review is found in Unsöld, *Vjschr. astr. Ges., Lpz.*, **78**, 213, 1943.

† Weisskopf, *Z. Phys.*, **75**, 278, 1932 and **77**, 398, 1932; Lenz, *Z. Phys.*, **80**, 423, 1933; Burkhardt, *Z. Phys.*, **115**, 592, 1940.

‡ Lindholm, *Ark. Mat. Astr. Fys.*, **28**, B, No. 3, 1942; van de Hulst, in course of publication.

§ E. Strömgren, *Stromgren. Festschr.*, 1940.

|| ten Bruggencate and Houtgast, *Z. Astrophys.*, **20**, 149, 1941.

¶ Woolley, *M.N.*, **94**, 631, 1934; *Observatory*, **57**, 345, 1934; Pannekoek, *M.N.*, **95**, 725, 1935; Strömgren, *Z. Astrophys.*, **10**, 237, 1935; Dempster, *Ap. J.*, **96**, 295, 1942.

another one, and a wave-length shift is observed.* A special case of this is the "interlocking", studied by Woolley. It was assumed that this effect would be absent in resonance lines, because the ground level is perfectly sharp. But Houtgast made plausible that a level disturbed by collisions must be considered as unsharp just as well as a naturally widened level. † Quantum mechanical calculations made by Zwaniken under the direction of my colleague Rosenfeld seem to confirm this view, but at the same time show the complications involved.

Having studied our volume element we proceed to the investigation of the solar atmosphere as a whole, and we have to consider how the interaction of all its volume elements explains the observed Fraunhofer spectrum. Our study of the properties of solar gases was actually atomic physics. The interaction of all the infinitely numerous layers is a typical astrophysical problem. These problems of transfer in immense atmospheres have a beauty of their own, which you only feel when you have worked in this field. The infinitely complicated paths of the radiation, with all their turns and bends and loops, are grasped by the power of mathematics into simple, impressive equations. The whole architecture of a stellar atmosphere may thus be studied in detail, with its temperature, pressure, ionization at each level.‡ In the last few years the work of Wildt and other scientists, culminating in that of Chandrasekhar and his collaborators, has furnished a most important basic contribution to this description of the Sun.§ If I could show you a sample of the solar gases, you would perceive that this mixture has a striking green-blue colour. It is due to the continuous absorption by negative hydrogen ions, of which there is only one against millions of ordinary hydrogen atoms. This gas, practically unknown on Earth, was discovered from the study of the solar spectrum; the calculation of its optical properties was carried out entirely by theory. Now for the first time it is possible to find the true relation between pressure and temperature in the successive layers. And this again is the basis for all calculations on the Fraunhofer line profiles.

There remains an uncertainty concerning the exact temperature distribution in the outer layers of the Sun. Part of the energy, absorbed by the Fraunhofer lines, is used in raising the temperature of the gas. This is the blanketing effect discovered by Milne. For its computation we ought to know how much energy is taken away from the solar spectrum by the whole of the Fraunhofer lines. Years ago I determined for this fraction 0·15. Mulders made a much more accurate measurement and found 0·08.‖ A few years ago however the French astronomers have shown that our energy curve has to be changed considerably in the ultra-violet; taking this into account, and repeating the computation, we come now to a coefficient 0·10. But the blanketing effect is more complicated than it was thought at first. It is different according to the height where the

* Eddington, *M.N.*, **89**, 633, 1929; Woolley, *M.N.*, **91**, 979, 1931 and **98**, 624, 1938; Pannekoek, *Proc. Acad. Sci. Amst.*, **34**, 1352, 1933; Weisskopf, *Z. Phys.*, **85**, 451, 1933; *Observatory*, **56**, 302, 1933; Spitzer, *M.N.*, **96**, 794, 1936 and *Ap. J.*, **99**, 107, 1944; Zanstra, *M.N.*, **101**, 273, 1941.

† Houtgast, *Variations in the profiles of strong Fraunhofer lines along a radius of the solar disk*, Dissertation, Utrecht, 1942.

‡ Strömgren, *Tables of model stellar Atmospheres*, Copenhagen, 1943.

§ *Ap. J.*, **104**, 430, 1946 and **104**, 446, 1946.

‖ Minnaert, *Z. Phys.*, **45**, 610, 1927; Mulders, *Aequivalente breedte van Fraunhofer-lynen*, Dissertation, Utrecht, 1933.

Fraunhofer lines originate and according to their mode of formation, either by scattering or by thermal absorption; and even according to their spacing along the spectrum. Another source of uncertainty is turbulence, which was shown by Plaskett to modify the general temperature distribution.

Fortunately for many problems concerning Fraunhofer lines these doubtful points are not essential. I would like to show you how easy and straightforward the calculation of the intensity becomes if it is applied to the important case of faint lines or to faint wings of strong lines. We are able to express the dip of the intensity in the line as a simple integral over the scattering and over the absorbing material, the contribution of each layer being weighted by a function of depth computed once for all *:

$$\frac{i_0 - i}{i_0} = \int_0^\infty \frac{\kappa_\nu}{\kappa} g_1(\tau) d\tau + \int_0^\infty \frac{\sigma_\nu}{\kappa} g_2(\tau) d\tau.$$

The integrals of $\frac{\kappa_\nu}{\kappa}$ and $\frac{\sigma_\nu}{\kappa}$ may be taken for practical purposes over the optical depth τ, but they are equivalent to integrals of κ_ν and σ_ν over the gas mass. So it is clear that the weighting functions simply express how effective are the scattering and selectively absorbing atoms at each depth: the highest layers are distinctly visible, whereas the deeper layers are veiled more and more and are finally lost in a luminous fog of continuous radiation. The computation of this weighting function can be made, independently of assumptions about the temperature distribution, by making use of the direct observational data on limb darkening. Making use of the analysis already carried out by Chalonge and Kourganoff †, this computation is easy and straightforward. We select wave-lengths where no influence of neighbouring Balmer lines is to be feared and we express the dip of the profile relatively to the continuous background. The weighting curves, thus computed for the centre of the Sun, expressed in units $E_{0\nu}$, are presented here as a practical help for the use of our atlas and for other work (Fig. 3). They are based on Abbot's measurements of 1906–08, those of the same author in 1913, and those of Raudenbusch in 1937. Several of the curves have been independently derived from two or three of these sets of data; it will be seen that the resulting differences, indicated by vertical bars, are small, and that the curves apparently are reliable to a fairly high degree. It is interesting to see how small is the difference between the effects of scattering and of thermal absorption, how the first is gradually getting more important the more we proceed to the red.

For a practical application of these formulae, it is necessary to divide once for all the functions $g(\tau)$ by $\kappa(\tau)$, and it is easier to integrate over $\bar{\tau}$ than over τ:

$$\frac{i - i_0}{i_0} = \int_0^\infty \kappa_\nu(\bar{\tau}) \, . \, h_1(\bar{\tau}) d\tau + \int_0^\infty \sigma_\nu(\bar{\tau}) \, . \, h_2(\tau) \, . \, d\bar{\tau}.$$

The discovery of the H^- absorption furnishes us with the data needed. As an illustration we assume the calculations of Barbier ‡ on the electron pressure and temperature, and use the variation of κ with wave-length as computed by

* Unsöld, *Z. Astrophys.*, **4**, 339, 1932 and **8**, 225, 1934; Minnaert, *Z. Astrophys.*, **12**, 313, 1936. A special case is studied by ten Bruggencate in *Naturwiss.*, **33**, 91, 1946; and more detailed calculations by the author are found in *B.A.N.*, **10**, 339, 1946.

† Chalonge and Kourganoff, *Ann. Astrophys.*, **9**, 69, 1946.

‡ Barbier, *Ann. Astrophys.*, **9**, 173, 1946.

Chandrasekhar. Fig. 4 now shows the weighting curves in their ultimate form, in units which are absolute and the same for all curves. Multiply the ordinates by $\sigma_\nu(\bar{\tau})$ and $\kappa_\nu(\bar{\tau})$, expressed per neutral hydrogen atom, integrate, and you get the dip of the line profile, expressed as a fraction of the continuous spectrum

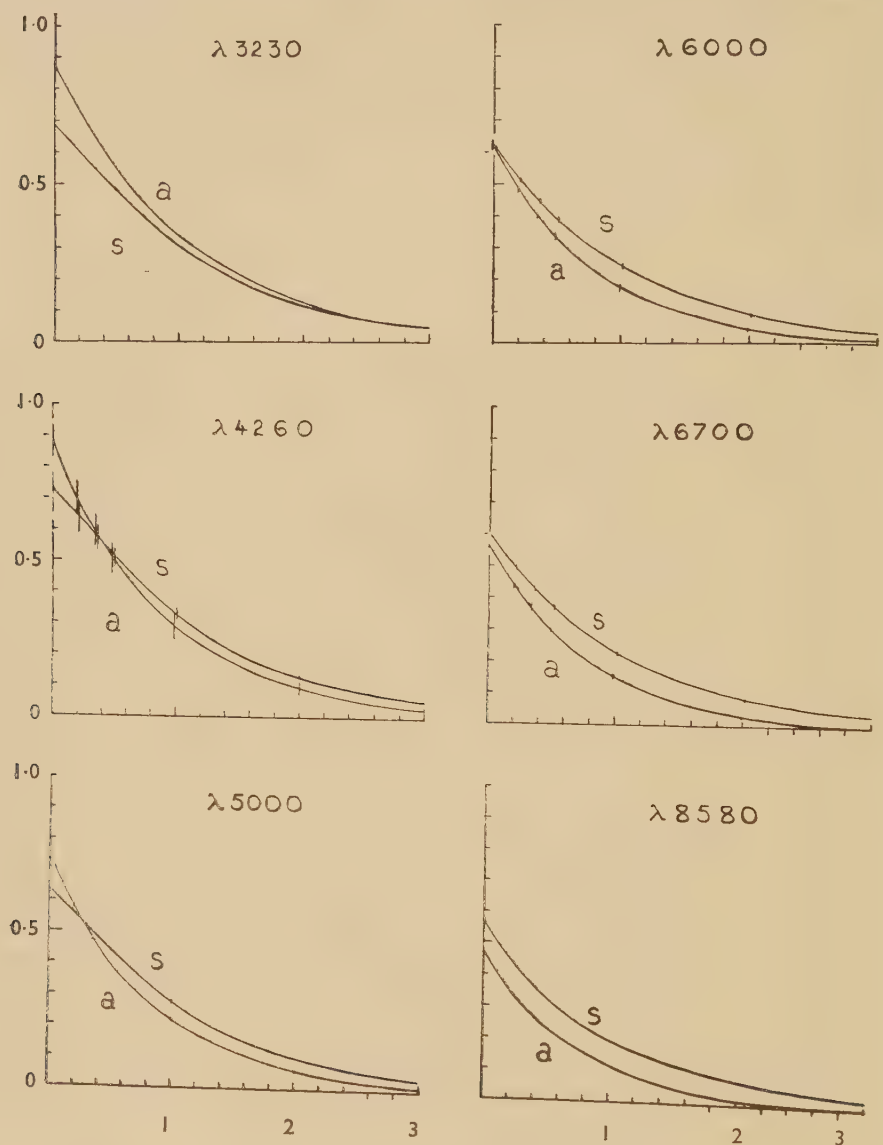


FIG. 3.—*Weighting functions* $g_1(\tau)$, $g_2(\tau)$.
Ordinates : g
Abscissae : τ

at the centre of the Sun. For such lines as H and K, for example, σ_ν is constant with depth and κ_ν is probably small, so the integral over our second weighting function is directly a measure for the dip. It is now still more apparent that only a small depth of the solar atmosphere contributes to the line formation; and this may be understood, because the effect of the rapidly increasing general

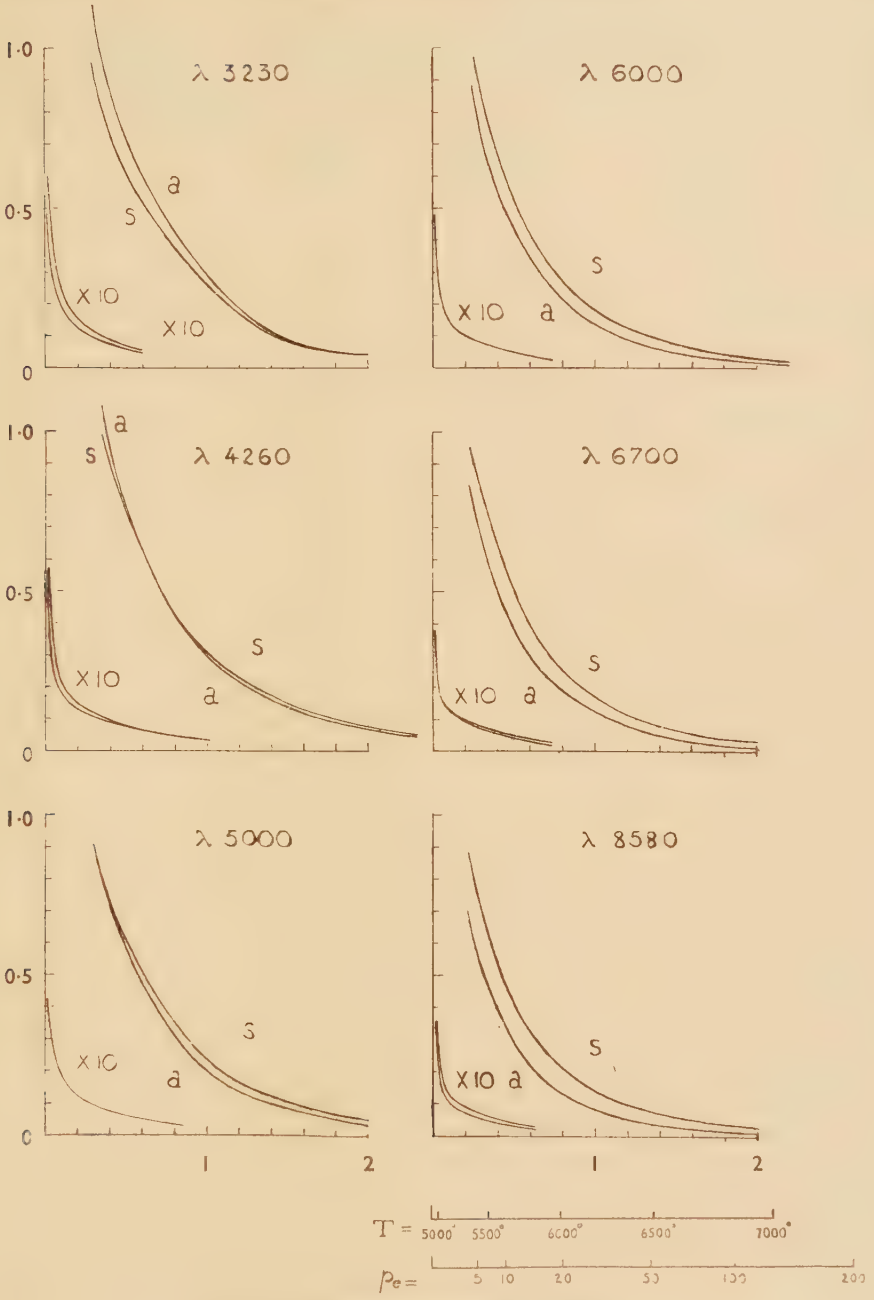


FIG. 4.—*Weighting functions* $h_1(\bar{\tau})$, $h(\bar{\tau})$.
Ordinates: $10^{-24}h$
Abscissae: $\bar{\tau}$

absorption has been brought into account. The effect of a given number of scattering or selectively absorbing atoms is now clearly seen to diminish towards the red, because the general absorption of H^- increases so that the deeper layers become invisible.

The theoretical calculation of profiles is less simple when the dip becomes more than about 15 per cent. Several simplified models have been proposed

as approximations; but for many investigations, as for example the centre to limb variations, they will not be sufficiently precise. Fortunately a detailed calculation is always possible without real difficulties, according to the method initiated by Pannekoek*: (1) We start from Eddington's equations:

$$\frac{dJ}{d\tau} = \left(1 + \frac{\kappa_\nu}{\kappa} + \frac{\sigma_\nu}{\kappa}\right)H,$$

$$\frac{dH}{d\tau} = J - E,$$

and make a table of the functions κ_ν/κ, σ_ν/κ, E, as derived from our present theory of the solar atmosphere. (2) We assume a certain value for H_0 and $J_0 = 1{\cdot}8H_0$ at the surface, and integrate by finite steps, proceeding toward greater depths. If the initial value has been selected right, J tends toward E, or, more precisely, toward $E + \frac{1}{3}d^2E/d\tau^2$. (3) Knowing E and J at each depth, the emergent intensity in any direction is obtained by a well-known general formula. Since the discovery of the H^- absorption has given a considerably modified but consistent picture of the solar atmosphere, and since the selective effects are better understood now than before, a systematic calculation as here explained would be very interesting. This has not yet been made.

I have tried to give you a bird's eye view on the development of profile measurements and on the present state of theory. I would like now to show the crucial points where the cooperation of observers and theorists will be most necessary and where great progress in the next years will have to be made. I think there are two such primarily important fields for future research, and it was for this purpose that I left these for the end of my lecture, in order to give them the full stress.

The first crucial point is the variation of Fraunhofer lines over the Sun's disk. In studying this, we investigate the influence of a new variable, which is the direction of observation relative to the solar surface. Actually we look at the same layers of the Sun, but literally speaking from a new point of view, and we are able to check whether we have pictured aright our object and its structure in space. In principle the method is remarkably similar to a parallax measurement: we shift the position of our eye, and at once the world around us shows depth and becomes three-dimensional.

In the very first period of solar research, after Fraunhofer but before Kirchhoff, the difference between the spectrum in the centre and near the limb had been already considered and had given the greatest trouble to astronomers. They pictured the Sun as a glowing solid body, surrounded by a selectively absorbing atmosphere, and they expected to see the Fraunhofer lines increase in strength toward the limb, just as the atmospheric lines of the Earth do toward sunset. But nothing of that kind happened. Already in 1836 Forbes had noticed during an annular eclipse that the Fraunhofer lines looked just the same as in the centre spectrum†: an interesting anticipation of similar eclipse observations with modern quantitative technique, made by me in 1929 and by Redman in 1943.‡ One hundred years ago this difficulty was considered so insuperable that scientists

* Pannekoek, *M.N.*, **91**, 139 and 519, 1930; and *Publ. astr. Inst. Univ. Amst.*, No. 4, 1935. The method was applied to the general case by Minnaert and Houtgast, *Z. Astrophys.*, **12**, 96, 1936.

† *Phil. Trans.*, **126**, 453, 1836.

‡ Minnaert, *B.A.N.*, **6**, 151, 1931; Redman, *M.N.*, **103**, 173, 1943.

were about to reject the whole idea of the solar origin of Fraunhofer lines! The question was solved at that moment in a very sound way—by forgetting about it; so overwhelming was the evidence of Kirchhoff and Bunsen's discoveries, which followed a few years later.

It is since the work of Schwarzschild that we have understood that things are not so simple as was thought in the early 'fifties. At the limb of the Sun we look less deep into the photosphere; if the selective gas is mixed all through the continuously absorbing and emitting material, the effective path of the light decreases because of the smaller effective depth at the same time that it increases because of the greater inclination. Moreover, there is an important contribution of scattered light at right-angles to the direction of vision. The result will depend on the position of the selective layer, on the ratio between thermal absorption and scattering, on the wave-length, and on the part of the Fraunhofer line which is studied. All such different possibilities make this a subject of great and absorbing interest. Plaskett and the Oxford astronomers have done very important work in this field.* They have tried to take as a starting-point the direct observations and to derive by stepwise induction the properties of the solar atmosphere. At Utrecht, Houtgast and I have followed a more deductive method, making models of the atmosphere, calculating how the line profiles should be in the escaping radiation, trying to get a close fit to the observations.† Both methods have their advantages and their danger. In the use of the inductive method, small errors of measurement may have an unexpectedly great effect on the calculated solar model. The danger of the deductive method is that one gets a more or less satisfactory agreement with a certain model, and does not look for other models equally acceptable. It is evident that both methods must supplement each other.

It is remarkable that up to now there is not a single well-studied case in which the centre to limb variations have been quantitatively explained on the basis of the classic model of the solar atmosphere. But we must admit that most calculations have been made in an over-simplified way, and that the recent determinations of the H^- absorption and their consequences for the structure of the solar atmosphere have not yet been included. Among the most interesting new assumptions which have been tried, there is the non-coherent scattering considered by Houtgast. He showed that variations in the profiles occur in a systematic way all over the disk, and not only at the extreme limb; and he could only account for them by assuming non-coherent scattering. Several other investigations are being made now at our observatory on the centre to limb variation; in the first place for the hydrogen lines, which have such a specific importance; and secondly for the faint lines, where the theory is remarkably simple. Our observers always compare the profiles on at least six points, at increasing distances from the centre.

From the combined efforts of the workers in this field we must be able to derive the *stratification* of the solar atmosphere: the distribution of its atoms and ions over the consecutive layers, but also the variation with depth of the damping and the thermal absorption. The conception of a "reversing layer"

* Plaskett, *M.N.*, **91**, 870, 1931, and **101**, 3, 1941; M. G. Adam, *M.N.*, **98**, 112, 1937, and **100**, 595, 1940; Evans, *M.N.*, **100**, 156, 1940, and **101**, 21, 1941.

† Minnaert and Houtgast, *Z. Astrophys.*, **12**, 81, 1936; Houtgast, *Z. Astrophys.*, **16**, 43, 1938, and Dissertation, Utrecht, 1942.

has definitely disappeared. We see the integrated effect of the selective gases up to optical depths of about 2, when looking at the centre of the Sun, and only up to 0·5, when looking at the limb.

The second crucial point where observation and theory ought to cooperate closely, is the individual comparison between the several spectral lines of an element. Up to now we have worked more or less statistically, but in future a detailed study will be necessary. Génard and I have made measurements on the P-D lines of magnesium, which then found a beautiful explanation by Unsöld. Righini, when working at Utrecht, compared the individual sodium lines to their predicted intensities. Allen and K. O. Wright showed that the iron and titanium lines in the Sun are in better agreement with laboratory intensities than with the elementary multiplet rules. Allen tried to find the line to line variations of the Stark effect in the solar iron spectrum.*

The use of a mean curve of growth is too crude a method for such researches, and should be replaced by assigning to each line the individual curve of growth to which it belongs. This is possible. For each element, we can determine the *concentration* from faint lines, which are hardly influenced by the damping. To these faint lines we compare each of the stronger lines, and are then able to derive their individual *damping* constants. In this way both factors may be separated.

It is especially in view of this whole programme of research that our atlas was published and that we are working out now our intensity catalogue.

I hope that this *aperçu*, limited to essential problems which almost all are related to the Utrecht investigations, will have convinced even the non-specialist astronomers here present that the spectrophotometry of Fraunhofer lines is a beautiful and fruitful field of investigation. It is often said that we use the stars as cosmic laboratories where matter is found in conditions unknown on Earth. Indeed, the curve of growth, the damping by phase disturbances, the non-coherent scattering, the van der Waals' forces, the existence of H^-, they all have been discovered in the Sun or they have found there their first extensive applications. The International Astronomical Union intends to publish a table of atomic transition probabilities, based on the Fraunhofer lines, on data of experimental physics, and on theoretical calculations. Such a table would be of the greatest importance for physics and for astrophysics. It will have become clear that these contributions from solar spectroscopy have only become possible after we are more or less able to discern the factors, determining the line profiles.

With respect to our knowledge of the Sun, the main object of all our work will be to reduce a great diversity of apparently incoherent facts into logical consequences of a few fundamental physical laws, the ultimate result for astronomy proper being : the relative abundances of the several elements. In this stage of our science we remember a sentence of Auguste Comte† in his *Cours de Philosophie Positive of* 1835 : " There are questions which for the human mind will remain for ever unknown, for example the composition of the celestial bodies ".

Once again we see that no one should try to restrict the grand sweep of science or of human progress.

* Minnaert and Génard, *Z. Astrophys.*, **10**, 377, 1935; Righini, *Z. Astrophys.*, **10**, 344, 1935; Allen, *M.N.*, **96**, 144, 1936, and **100**, 4, 1940; K. O. Wright, *Ap. J.*, **99**, 249, 1944.

† *Cours de philosophie positive*, **2**, 8, Paris, 1835.

A PROPOSED "MIRROR TRANSIT-CIRCLE"

R. d'E. Atkinson

(Received 1946 September 6)

Summary

A fundamental transit-circle is proposed, in which the usual moving telescope is replaced by a moving plane mirror, with its normal sweeping out the meridian. Observation is through either one of two horizontal telescopes, lying in the meridian in the positions at present occupied by the two collimators; these telescopes also act as collimators when required.

A number of advantages of this method are listed, and also a few disadvantages; the latter, however, appear relatively unimportant.

It is shown that if the rear face of the mirror is roof-shaped, half of it being inclined laterally towards the east by about 15°, and half towards the west, extra-meridian observations of declination can be used to obtain fundamental azimuths in a period of about four hours instead of twelve.

Equations are developed for the position of the point observed, and the tilt of the star's path at apparent transit, for the case when this mirror-tilt cannot be treated as a small quantity; it appears that the transit does in fact occur nearly horizontally over a considerable range of declinations fairly near the zenith, and that the accuracy of azimuth determinations should be satisfactory.

The general operation of the instrument is discussed, including the determination of the collimation; the collimation error cannot be represented in the form $c \sec \delta$, and the necessary expression is derived from the equations already obtained, by making the approximations appropriate for very small tilts. The instrument has a considerable number of self-checks which are not possible with the ordinary type, and allows a particularly fundamental method of obtaining both its pivot errors and the personal equation of the observer.

The mounting of the mirror is discussed, and it is suggested that the final form of the instrument would be with both telescopes broken towards the east, so that the two observing positions are close together, with a permanent pivot-telescope between them; that the focal lengths might be about 30 feet with the magnification not much changed from the present figure; that the telescopes could be very elaborately jacketed; and that if desired some quite different and much more bulky arrangement than a screw could be used for the micrometer drives.

In the course of some recent work at Aberdeen, Maryland, U.S.A., it occurred to me that many of the serious difficulties which still beset the transit-circle ought to be diminished, or even eliminated altogether, if one substituted a moving plane mirror for the moving telescope. The mirror would be mounted with its plane parallel to the existing rotation axis, so that its normal would sweep out the meridian, and one would observe stars reflected in it, through one or the other of two fixed telescopes, situated in the meridian; these might very well in fact be the existing collimators, if they were fitted with suitable driven-micrometer eyepieces.

The advantages to be expected from such an arrangement are really very considerable. We may note, in particular, the following:—

1. There will be practically complete freedom from variable flexure (droop) due to varying weight stresses. The telescopes will not now move at all, and the

mirror, if in fact it were to bend perceptibly, could cause no first-order effects except an astigmatism; in practice, the flexure of any reasonably thick mirror would be quite inappreciable. It would, of course, have to be very well mounted, and this question is considered below.

2. There will be practically complete freedom from telescope-flexure due to temperature differences. Although such distortions can be determined by repeated collimation and nadir observations, it is very time-consuming to have to do this at all frequently; but the effects can certainly be too large to ignore by modern standards, and may also vary rapidly. Fixed telescopes can be shielded to any desired extent, including the use of circulating oil-baths etc. if necessary.

3. The weight relations in the micrometer will be constant. There is evidence that the shifting relations inseparable from a moving telescope can cause variable collimation and/or nadir depending on the altitude. With a fixed telescope, the micrometer can be designed with special reference to the one direction in which gravity will act.

4. The observing position will be constant, including that for observations of levels and nadir. This should make the personal equation more constant; moreover, since the position can be made a permanently convenient one, it should diminish fatigue.

5. It would be easy to arrange the two driven micrometers so that one acted as a personal-equation machine for the other. Observers would thus determine their own personal equation with the actual instrument, and in the actual observing position, which they also used for the observations.

6. Very much shorter piers can be used; they need only be tall enough to accommodate the circles. This will greatly decrease the changes of level caused (e.g. when the wind changes) by changing temperature differences between the piers; it will decrease any azimuth changes that may be caused by one pier inclining more north than the other (for example if its south face is warmed); and if it is desired to observe stars after reflection in a mercury pool, the size (or the range of displacement) of the pool can be kept small.

7. A very much longer focal length can be used, with or without any corresponding increase in the magnification. Since the aperture would probably not be increased very much, the lens corrections would be more satisfactory, especially those for spherical aberration, coma, and curvature. In addition, backlash, screw errors, etc., would become less important. The magnification should probably not, in fact, be much increased, so that the eyepiece focal length would also be made considerably greater; this would give more room for reversing prisms, etc., and less trouble with the parallax, or focus difference, between the two moving wires; moreover, the wires themselves would appear thinner.

8. Any desired modifications to the micrometer system can be planned without having to keep a strict eye on considerations of weight and space. For example, the position of the wire could control a rotating mirror, and a fixed scale could be photographed every second by reflection in the mirror; such an arrangement would practically eliminate backlash.

9. The aperture can also be increased without difficulty, if this is in fact desired. A very small increase would take care of the loss of light due to the single reflection in the mirror.

10. The connections for the micrometer drive, field illumination, chronograph, etc., will be simplified; neither slip-rings nor flexible cables will be needed any more, and all leads can be permanently run with fixed cable.

11. Reversing will be simplified, since no leads at all have to be disconnected; moreover, the body to be reversed will be much smaller.

12. Nadir and levels can be taken in either collimator. The differences between the two determinations must check with the division errors, in the one case, and with the pivot errors in the other.

13. A considerable range of near-zenith stars can also be observed in either collimator. The differences must again check with the division or pivot errors.

14. A particularly fundamental method of obtaining the pivot errors themselves can be provided by squaring, figuring and silvering a small patch on one or both edge faces of the mirror block, and taking autocollimation on this through the corresponding pivot. If flexure of the axis or movement of the mirror contributes in any way to the effective pivot errors, its contribution will automatically be exactly included in a determination made in this way.

15. As will be shown below, it is in principle possible to give the mirror a tilted rear face, making extra-meridian observations possible, and to use declinations obtained in this way so as to determine absolute azimuths fundamentally in a period of about three or four hours instead of twelve. This would be a major advantage if the practical difficulties can be overcome. Since the necessary observations do not involve any sub-pole measurements, it would thus become possible to do fundamental astronomy in the tropics without relying on low-altitude measurements, and indeed the method is in principle most sensitive (or alternatively, most rapid) in low latitudes. It has, however, some drawbacks.

In view of these apparent advantages, I was surprised to learn, through discussion with various astronomers at Aberdeen and elsewhere, that the idea of a mirror transit-circle in any form seemed unfamiliar. On my return here, however, Mr Martin pointed out that it is not in fact new; essentially the same proposal was made by Turner in 1894*, but it has apparently lain disregarded since then. It is true that Turner did not propose to use the collimators as observing telescopes; he proposed two or more special telescopes, to be arranged so that the observer might be in a comfortable warmed room. This does not appear advisable, judged by modern standards of accuracy, since it would introduce an awkward temperature gradient into the optical path; moreover, the collimators are indispensable in any case, and it is clearly desirable to have as few independent components as possible. Turner also did not discuss the actual process of collimation, and did not give any of the necessary formulae; nevertheless, the idea was in its essence proposed by him, and it does appear surprising that in spite of his prestige, and the many apparent advantages (of which he listed a considerable number himself), no attempt has yet been made to put the proposal into practice.

A mirror transit-circle has of course some disadvantages of its own, and we will now consider these briefly. On the whole, they appear insignificant by comparison with the advantages. The following seem to be the principal ones:

1. An error in the circle-readings, or in the determination of the division-errors of the circle, will have twice as much effect on the declinations as with a

* H. H. Turner, *M.N.*, **54**, 412, 1894.

telescope. However, these errors are not at present the limiting factor even for the internal consistency, and it appears that the expected improvement in systematic accuracy should hardly be impaired by doubling them.

2. An error in the determination of the pivot errors will have an increased (but never quite doubled) effect on the right ascensions ; the determination of pivot errors has not always been sufficiently thorough in the past, but it can in fact be made with more accuracy than that with which the star can be observed, so that here also doubling the uncertainty could perhaps be tolerated.

3. There will be a small reflection-loss of light; but, as already mentioned, this can be taken up by a very slight increase in aperture, since weight is no longer a consideration.

4. There will be a lower limit to the altitudes observable; below this, the cell of the telescope lens will interfere with the incoming beam. The lowest altitude, A, free from interference is given by $\tan A = d/D$, where d is the diameter of the objective (more accurately, the mean of the clear lens-diameter and the outside cell-diameter), and D is the distance from the bottom edge of the lens to the (nearest point of the) mirror. It is clear that even if d were as much as 25 cm., altitudes as low as 6° could be observed with the two collimator lenses separated by only 5 m.; this is a smaller separation than is often found, and it may be questioned whether astronomical measurements at lower altitudes than 6° have any general reliability.

5. It is not immediately clear how azimuth-marks can be viewed, since in general it would be undesirable to place them at 6° elevation, and the observer's body would now be in the way at 6° depression. This question is considered further below.

6. The operation of determining collimation must still include sighting the one collimator on the other. It would be undesirable to have the mirror so far away from the rotation axis that when it was laid horizontally there was a clear view through, above or below it, and there thus seems no alternative to raising it a few inches every time collimation is taken. However, it would be easy to design the weight-relieving apparatus so that this short lift was carried out by it rather than by the reversing apparatus ; and it does not appear likely that any difficulty or inconstancy would then result from such repeated raisings. If it does, it would presumably do so with a telescope-transit also, but would be less readily noticed.

It may well be that there remain other disadvantages which have been overlooked; but as far as the above survey goes, there can be little doubt that the proposed instrument would be considerably more satisfactory, in respect of systematic errors and also in respect of several accidental ones, than the existing type, which has in fact suffered practically no fundamental improvement, except the introduction of the impersonal micrometer, for over a century.

We will now consider the operation of the instrument. The process of observing a star will be exactly as with a telescope-transit, except only that the observer's own line of sight is always horizontal. The choice between observing in the north collimator or the south one, for all stars within (say) 20° of the zenith, is new, and one would in fact arrange a suitable amount of alternation, both for the programme of any one night and for the total schedule for any one star. In setting on a star, the circle must be set with only half the previously allowable tolerance to bring it into the field of view ; and in moving from one star to another,

without change of collimator, the circle moves only half as far as before. The level error and nadir may of course be obtained by reflection in a mercury pool as before, by turning the mirror so that its normal is 45° away from vertically downwards. In the absence of collimation error, if we trace a central ray backwards from telescope to mirror to pool, it will hit the pool vertically if there is no level error; and if there is a level error it will hit the pool at an angle differing from vertical by once (not twice) this error; thus the numerical value of the level is the same, in terms of micrometer readings, as it would be for a telescope-transit of equal focal length. Difference in level on reversal of the instrument will give the difference in pivot-diameter, as at present. An azimuth-error of the instrument as a whole will also appear from the star observations in exactly the same way as before, i.e. without any factor of two or other correction. Collimation, however, is radically different from collimation with a telescope-transit, and we must study this in detail.

We have here in fact an entire additional parameter, beyond those which occur with a telescope-transit. There the collimation was simply the angle, at the lens, between the line through the micrometer-wire in what may be called the "position of the standard contact" and the line perpendicular to the rotation axis; with the mirror-transit we shall have to consider in addition to this angle the angle between the plane of the mirror and the rotation axis. Both angles are of course readily obtained, the procedure being as follows.

Using a Bohnenberger autocollimating eyepiece, one first obtains for each collimator the setting that gives autocollimation in R.A. in the mirror. Let these readings be N_0 and S_0 for the north and south collimators respectively. Then if we measure azimuths from the direction perpendicular to the rotation axis (clockwise from "south"), and if the angle between mirror and axis is β_1 (positive if the normal to the mirror is tipped west of the "meridian" plane), the north collimator is pointing in an azimuth $-\beta_1$, and the south one in an azimuth $\pi+\beta_1$, when the readings are N_0 and S_0. Now raise the mirror, and sight the north collimator on the south one when the south one reads S_0. Let the north setting for this case be N_1; for positive β_1, N_1 will be greater than N_0 if the N micrometer readings increase as the wire moves east. At the setting N_1, the north collimator is evidently pointing in an azimuth $\pi+(\pi+\beta_1)=\beta_1$, and the setting which gives an "azimuth" of zero, i.e. perpendicular to the rotation axis, is $(N_1+N_0)/2$. If the values of N are in mm., and if the focal length is F_N mm., the value of β_1 in seconds of arc is

$$\beta_1 = 206265\,\frac{N_1-N_0}{2F_N}. \qquad (1)$$

Similarly, the setting of the south collimator which gives an "azimuth" of π, i.e. perpendicular to the rotation axis, is $(S_1+S_0)/2$ if S_1 is the setting of the S collimator when it is sighted on the N one reading N_0; and

$$-\beta_1 = 206265\,\frac{S_1-S_0}{2F_S} \qquad (2)$$

if the S micrometer readings increase as the wire moves west. The two determinations of β_1 should of course agree, and a further obvious check is that when the north collimator reads $(N_1+N_0)/2$ and the south one $(S_1+S_0)/2$, they must appear set on each other, whether or not $F_N=F_S$.

In the analysis which follows, we will also consider that the collimators are horizontal; this can obviously be arranged by comparing nadir readings with autocollimation readings, adjusting the zero of the Z.D. micrometer until the circle readings differ by 45° in the two positions. When both collimators have been adjusted in this way, if the mirror is again raised, the horizontal wires should appear set on each other, as well as the vertical ones.

We may note in passing that these operations, which will in fact form part of the standard routine, give us three independent checks, to which we may add a fourth from the comparison of the level as obtained in both collimators. Of these four, two have their counterparts in similar checks with a telescope-transit (one is used for flexure determinations); but the two independent nadir determinations, and two independent level-determinations, are peculiar to the mirror-transit.

We will now consider the geometrical relations which hold, and the corrections which must be applied, when β_1 is not zero; in view of the possible application to a mirror with a deliberately tilted rear face, we will develop the analysis rigorously, i.e. without at first making any of the approximations that are admissible when β_1 is very small. For the front of the mirror, it should in fact be possible to reduce the tilt to a very few seconds, and an analysis restricted to the first power of β_1 will be amply accurate; but the tilts required for the back of the mirror may be 10 or 15 degrees.

We will continue, for the present, to speak of "azimuths" as though the instrument as a whole were correctly oriented, i.e. we shall measure them from the direction perpendicular to the rotation axis. "North", "west", "meridian", etc., are to be understood in this sense.

Consider now spherical polar coordinates with the poles on the observer's horizon, at the east and west points. Let "latitudes", β, on this system be measured positive to the west, from the meridian, and "longitudes", λ, positive up from the south horizon, so long as we are concerned with the south collimator. Then if the mirror is rotated through a "longitude" angle λ_1, up from the south autocollimation position, the coordinates of the normal are (λ_1, β_1). Let the coordinates of the star which is central in the field of view in this position be (λ_2, β_2). Then if θ is the angle from the south point to (λ_1, β_1), it is also the angle from (λ_1, β_1) to (λ_2, β_2); moreover, these three points lie on a great circle. In Fig. 1, O is the star; N, W, S, the north, west, and south points of the horizon; and M the normal to the mirror. Then from the triangle MWS we have

$$\cos\theta = \cos\lambda_1 \cos\beta_1 \tag{3}$$

$$\sin\theta \cos\psi = \sin\lambda_1 \cos\beta_1 \tag{4}$$

$$\sin\theta \sin\psi = \sin\beta_1 \tag{5}$$

where the angles are as marked in the figure.

From the triangle OWS we have

$$\sin\beta_2 = \sin 2\theta \sin\psi \tag{6}$$

$$\sin\lambda_2 \cos\beta_2 = \sin 2\theta \cos\psi \tag{7}$$

$$\cos\lambda_2 \cos\beta_2 = \cos 2\theta = 2\cos^2\lambda_1 \cos^2\beta_1 - 1 \tag{8}$$

by applying (3).

From (3), (5) and (6),

$$\sin\beta_2 = \cos\lambda_1 \sin 2\beta_1 \tag{9}$$

and from (3), (4) and (7),

$$\sin \lambda_2 \cos \beta_2 = \sin 2\lambda_1 \cos^2 \beta_1. \tag{10}$$

Any two of the equations (8), (9) and (10) give the coordinates of the star in terms of the mirror angles λ_1 and β_1.

The track given by these equations, with β_1 constant and λ_1 variable, is not a small circle. It evidently passes through $\lambda_2 = 0$, $\beta_2 = 2\beta_1$, and through $\lambda_2 = \pi$,

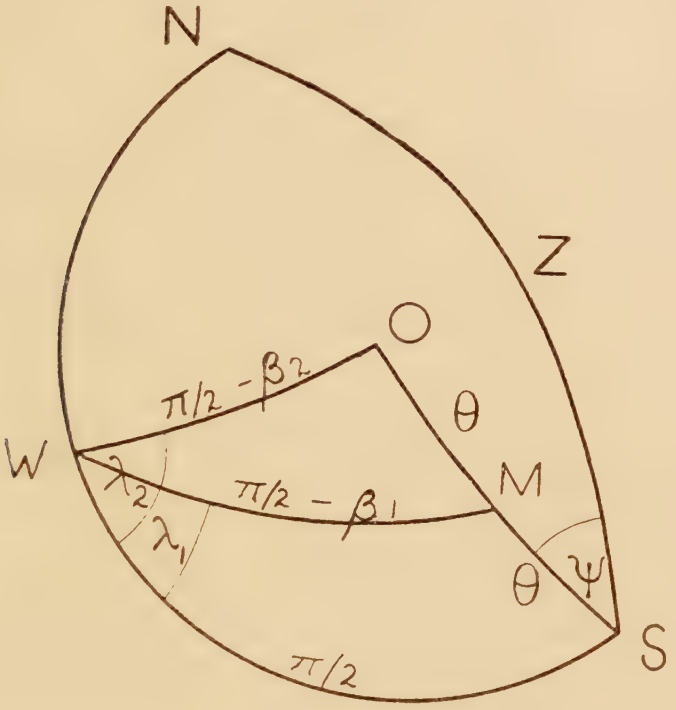


Fig. 1.

$\beta_2 = 0$; any small circle through these two points must culminate at an azimuth $\pi/2 + \beta_1$, but the actual track culminates at an azimuth A given by

$$\cos A \sqrt{(1 + \cos^2 \beta_1)} = -\sin \beta_1$$

instead of

$$\cos A = -\sin \beta_1.$$

In fact the track still departs from a small circle by quantities of the first order in β_1 when β_1 is small. Thus it cannot even in that case be adequately represented by any combination of *constant* fictitious level, azimuth, and collimation errors.

The hour angle H and declination δ of a star at "transit" may readily be obtained for any β_1, for the case in which the azimuth error of the instrument as a whole is zero. In Fig. 2, P is the celestial pole, and the other points have the same significance as before. From the triangle POS we have, if ϕ is the latitude,

$$\sin \delta = -\cos 2\theta \cos \phi + \sin 2\theta \sin \phi \cos \psi \tag{11}$$

$$\sin H \cos \delta = \sin 2\theta \sin \psi \tag{12}$$

$$\cos H \cos \delta = \cos 2\theta \sin \phi + \sin 2\theta \cos \phi \cos \psi \tag{13}$$

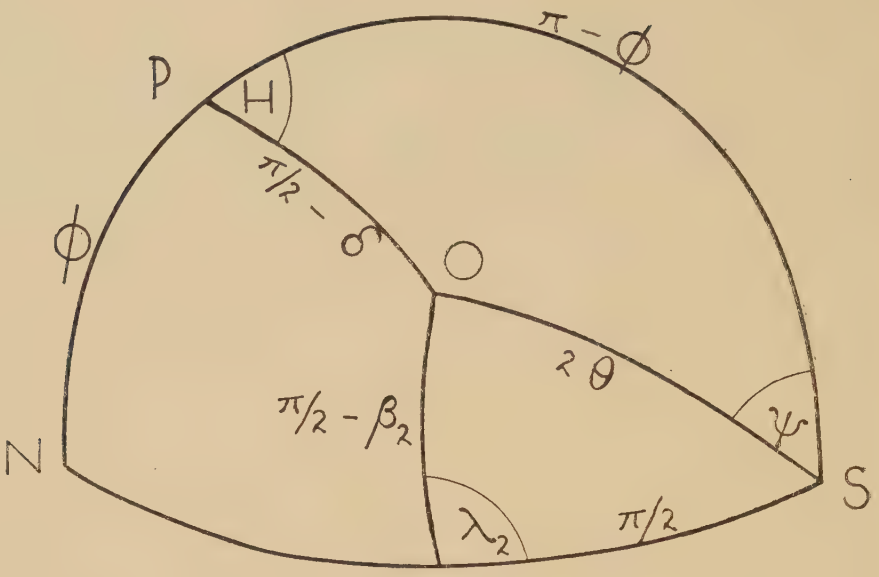


FIG. 2.

and thus from (11), (8), (7) and (10),

$$\sin\delta = \sin^2\beta_1\cos\phi - \cos^2\beta_1\cos(2\lambda_1+\phi); \tag{14}$$

from (12), (6) and (9),

$$\sin H\cos\delta = \cos\lambda_1\sin 2\beta_1; \tag{15}$$

and from (13), (8), (7) and (10),

$$\cos H\cos\delta = \cos^2\beta_1\sin(2\lambda_1+\phi) - \sin^2\beta_1\sin\phi. \tag{16}$$

Any two of the equations (14), (15), and (16) give the hour angle and declination in terms of the mirror angles and the latitude.

The largest value of δ that can be observed, δ_1 say, is evidently given by the value of λ_1 which satisfies the equation

$$\frac{d\delta}{d\lambda_1} = 0 = 2\sec.\delta_1\cos^2\beta_1\sin(2\lambda_1+\phi)$$

from which $2\lambda_1+\phi=\pi$, and so from (14)

$$\sin\delta_1 = \sin^2\beta_1\cos\phi + \cos^2\beta_1 \tag{17}$$

which is naturally less than unity unless either $\beta_1=0$ or $\phi=0$. Since at the maximum $\cos\lambda_1 = \sin\frac{1}{2}\theta$, the corresponding hour angle H_1 is given from (15) and (16) by

$$\cot H_1 = -\tan\beta_1\cos\tfrac{1}{2}\theta. \tag{18}$$

When β_1 differs considerably from zero, the stars will in general transit obliquely. If the true azimuth and level of the instrument are both reasonably close to zero, the apparent direction of transit may be obtained as follows. The angle ψ is the angle between the vertical plane through the collimator axis and the plane through the mirror-normal and the collimator axis; if the normal, i. e. the plane SMO, were visible through the telescope, it would be seen tilted at this angle to the vertical. Any star moving along the arc SMO on the sphere produces a reflected ray still moving along this arc at the S point; i. e. this direction of motion

in the sky gives an apparent direction of motion at the angle ψ in the telescope. Relative directions around O in the sky are reproduced without distortion in the telescope field, and in particular, as appears from Fig. 3, the image of the hour circle through O appears as a line at the angle η from the ψ direction, where η is equal to POS. Thus the image of the small circle of constant declination through O makes an angle $\eta+\psi+\pi/2$ with the vertical, or $\eta+\psi$ with the horizontal. Now from the triangle POS we have

$$-\cos\phi = \sin\delta\cos 2\theta + \cos\delta\sin 2\theta\cos\eta \tag{19}$$

and

$$\sin\eta = \frac{\sin\phi\sin H}{\sin 2\theta}. \tag{20}$$

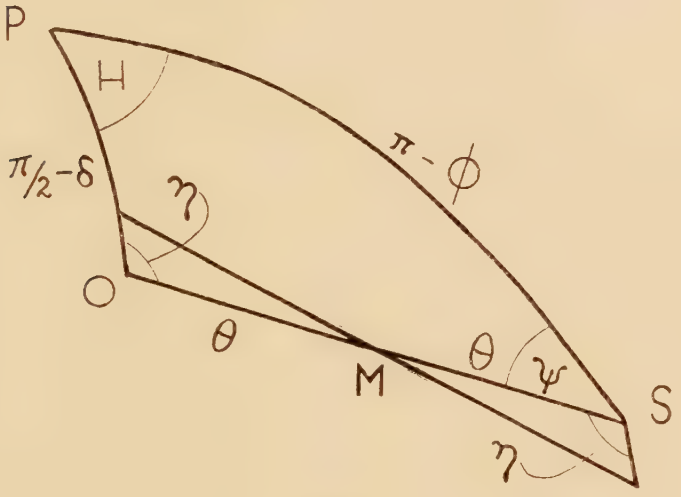


Fig. 3.

From (4) and (5) we have

$$\cot\psi = \sin\lambda_1\cot\beta_1 \tag{21}$$

and from (19) and (20)

$$\cot\eta = -\frac{\cos\phi+\sin\delta\cos 2\theta}{\sin\phi\sin H\cos\delta} \tag{22}$$

from which, using (14), (8) and (15), we obtain after some tedious reductions

$$\cot(\eta+\psi) = \sin\lambda_1\cot\beta_1 + \sec(\lambda_1+\phi)\frac{\sin\phi}{\sin 2\beta_1}. \tag{23}$$

Thus $\eta+\psi$, the angle from the horizontal at which a star appears to transit, can be tabulated in terms of λ_1, for any given values of ϕ and β_1.

An alternative formula for the tilt may be obtained by using one of the standard expressions for spherical excess. Since, in the triangle POS,

$$\eta+\psi+H = \pi+E \tag{24}$$

if E is the spherical excess, and since we can just as well consider either $\pi-(\eta+\psi)$ or $\eta+\psi$ to be the tilt, we may say that the tilt is $H-E$. We have then

$$\frac{\tan\frac{1}{2}(H-E)}{\tan\frac{1}{2}H} = \frac{\cos\frac{1}{2}\left(\pi-\phi+\frac{\pi}{2}-\delta\right)}{\cos\frac{1}{2}\left(\pi-\phi-\frac{\pi}{2}+\delta\right)} \tag{25}$$

or

$$\tan \tfrac{1}{2}(H-E) = \frac{\sin \tfrac{1}{2}\left(\phi+\delta-\frac{\pi}{2}\right)}{\sin \tfrac{1}{2}\left(\phi-\delta+\frac{\pi}{2}\right)} \tan \tfrac{1}{2} H. \qquad (26)$$

The equivalence of the expressions (23) and (26) appears a little cumbrous to demonstrate trigonometrically; it can of course be verified arithmetically for particular cases without difficulty, and this provides, in fact, a useful check on the computations of H and δ.

It is evident from (23) that $\eta+\psi$ will be zero (or π), i.e. the star will after all transit horizontally, if $\lambda_1 = \pi/2 - \phi$; in this case, the normal to the mirror points towards the equator, whatever the tilt β_1 may be. It is then easily verified also that $\delta = \pi/2 - \phi = \lambda_1$, and $H = 2\beta_1$, whatever the value of β_1; provided, of course, that δ does not exceed the limit given by (17), in which case no horizontal transit occurs. For moderate latitudes (say between 35° and 55°) the stars in question culminate not far from the zenith; these are, fortunately, quite suitable stars for azimuth determinations by this method.

The following table, calculated for latitude 50° 52′ 20″, shows for various β_1 the tilt T, hour angle H, and declination δ, corresponding to various circle-settings λ_1. For small λ_1, the rate of change of tilt with λ_1 is nearly constant and nearly proportional to β_1; by the time the tilt has reached zero, its rate of change has begun to rise a little, but linear interpolation will still give a good first estimate. H is comparable with $2\beta_1$, from $\lambda_1 = 10°$ up to a little beyond $\eta+\psi = 0$, and δ changes by roughly twice the change in λ_1, in this same range.

If now we observe a star at about the declination at which the transits are horizontal, it will be possible to follow it with the Z.D. micrometer, and to record this micrometer's readings at the instants when the star passes various fixed vertical wires; the collimations of these wires being known, the Z.D. micrometer reading can be obtained for the instant when the star passed the fictitious wire of zero collimation. Since the movement of the micrometer will be very slow, the various instants do not have to be obtained with high accuracy. The declination corresponding to the circle-reading is obtained from (14); and the Z.D. micrometer reading, reckoned from the zero already established, and reduced to seconds of arc, is directly added to this declination. (Strictly, the Z.D. micrometer reading should first be multiplied by the cosine of the tilt.) Thus the declination can be obtained in the tilted rear face of the mirror, just as it can in the front, and with equal accuracy.

Now let us suppose that the entire instrument has a small azimuth error a; this will be much too small to affect the tilt itself visibly, but it will affect the apparent declination. We may study the question as though the mirror had been exactly set on the star, i.e. λ_1 was so skilfully chosen that the Z.D. micrometer reading at "transit" was exactly zero. Then there are two points in the sky that we are concerned with; one is the point defined by λ_1 on the assumption that $a = 0$, and the other is the actual point observed, with the same λ_1 but with the azimuth error a. It is clear that these two points have the same zenith distance z, and that their azimuths differ by a. The length of the arc joining them is thus $a \sin z$, and resolving this in the directions of increasing declination and increasing hour angle we readily obtain from Fig. 4

TABLE I

β_1		λ_1	0°	10	20°	30°	39° 07′ 40″	40°	50°	60°	70°
5°	T		− 8 02′	− 5 07′	− 3° 16′	− 1° 41′	0	+ 0° 12′	+ 3° 50′	+ 20° 45′	+ 153° 50′
	H		$0^h 51{\cdot}2^m$	$0^h 41{\cdot}6^m$	$0^h 37{\cdot}6^m$	$0^h 37{\cdot}1^m$	$0^h 40{\cdot}0^m$	$0^h 40{\cdot}5^m$	$0^h 52{\cdot}7^m$	$1^h 59{\cdot}2^m$	$10^h 51{\cdot}6^m$
	δ		−38° 25′	−18° 41′	+ 1° 08′	+ 21° 00′	+39° 08′	+40° 52′	+ 60° 39′	+79° 57′	+ 78° 21′
10°	T		−15° 33′	−10° 03′	− 6° 26′	− 3° 20′	0	+ 0° 24′	+ 7° 25′	+ 34° 31′	+ 132° 16′
	H		$1^h 40{\cdot}5^m$	$1^h 22{\cdot}7^m$	$1^h 15{\cdot}0^m$	$1^h 14{\cdot}2^m$	$1^h 20{\cdot}0^m$	$1^h 21{\cdot}0^m$	$1^h 44{\cdot}4^m$	$3^h 30{\cdot}7^m$	$10^h 01{\cdot}8^m$
	δ		−36° 22′	−17° 23′	+ 1° 56′	+ 21° 23′	+ 39° 08′	+ 40° 49′	+ 60° 01′	+ 77° 35′	+ 76° 17′
15°	T		−22 08′	−14 37′	− 9° 27′	− 4° 54′	0	+ 0° 35′	+ 10° 33′	+ 41° 42′	+ 115° 48′
	H		$2^h 26{\cdot}6^m$	$2^h 02{\cdot}8^m$	$1^h 52{\cdot}3^m$	$1^h 51{\cdot}4^m$	$2^h 00{\cdot}0^m$	$2^h 01{\cdot}5^m$	$2^h 34{\cdot}5^m$	$4^h 35{\cdot}9^m$	$9^h 32{\cdot}5^m$
	δ		−33 08′	−15 17′	+ 3° 14′	22° 00′	+39° 08′	+40° 45′	+59° 01′	+ 74° 28′	+ 73° 27′
20°	T		−27 36′	−18° 41′	−12° 12′	− 6° 21′	0	+ 0° 44′	+ 13° 07′	+ 44° 46′	+ 102° 57′
	H		$3^h 09{\cdot}0^m$	$2^h 41{\cdot}6^m$	$2^h 29{\cdot}3^m$	$2^h 28{\cdot}7^m$	$2^h 40{\cdot}0^m$	$2^h 41{\cdot}9^m$	$3^h 22{\cdot}5^m$	$5^h 25{\cdot}1^m$	$9^h 17{\cdot}9^m$
	δ		−28° 55′	−12° 27′	+ 5° 00′	+ 22° 51′	+39° 08′	+40° 40′	+ 57° 41′	+71° 01′	+ 70° 13′

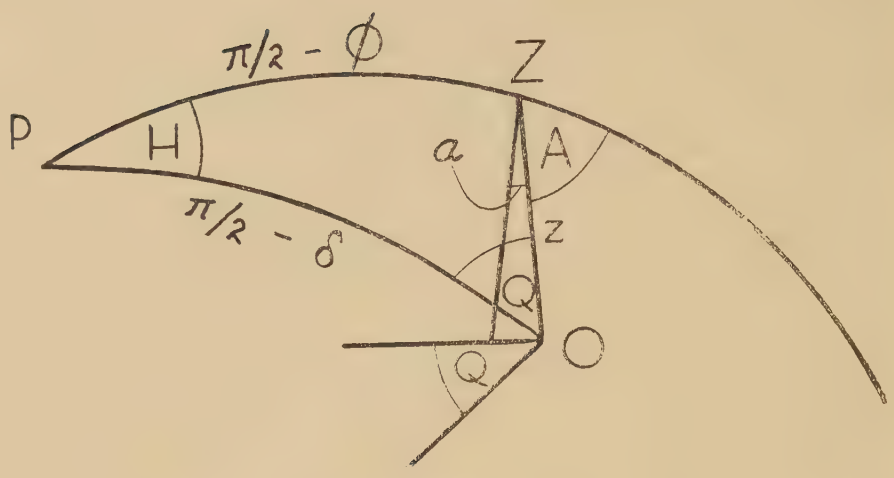


FIG. 4.

$$d\delta = a \sin z \sin Q, \tag{27}$$

$$\cos\delta \, . \, dH = a \sin z \cos Q, \tag{28}$$

where Q is the parallactic angle. We have

$$\sin z \sin Q = \sin H \cos\phi, \tag{29}$$

$$\sin z \cos Q = \sin\phi \cos\delta - \cos\phi \sin\delta \cos H, \tag{30}$$

so that

$$d\delta = a \sin H \cos\phi, \tag{31}$$

$$dH = a \sec\delta [\sin(\phi - \delta) + \cos\phi \sin\delta (1 - \cos H)]. \tag{32}$$

If we pick a value of β_1 that makes H about 2 hours, we have, for $\phi \approx 50^\circ$

$$d\delta \approx a/3.$$

It would appear desirable in fact to give the back of the mirror a roof shape, with half of it tilted east and half west. The apparent difference in declination if the same star were observed in both extra-meridian positions, i.e. two hours before and two hours after meridian passage, would then be $2a/3$ approximately, so that each star observed in this way would give a fundamental azimuth with an accuracy approaching quite reasonably close to that with which the difference of the two closely-equal "declinations" could be determined. Refraction, and most of the other likely sources of error, would cancel out in general. It is true that only half the effective aperture can be used in this way, but even so a large number of stars will be bright enough. The diffraction pattern will also in principle be a little unsymmetrical; but the asymmetry will cancel out as between the two observations. From Table I we see that, for the latitude in question and taking $H = 2$ hours (i.e. $\beta_1 = 15^\circ$), the angle of tilt, in the neighbourhood of zero tilt, changes by about 1° for each $1{\cdot}7^\circ$ of declination change; if we estimate that the method would be practicable so long as the tilt does not exceed $\pm 3^\circ$, we can make use of the belt of stars bounded by $\delta = \pi/2 - \phi \pm 5^\circ$, i.e. a ten-degree wide belt. This will clearly contain ample stars; however, it would be necessary to discover by trial whether a 3° tilt could in practice be accepted. Extra-meridian observations have of course often been made in the past, but perhaps not quite with the accuracy we are now demanding.

We may also enquire as to the declination at which the stars transit nearly vertically, since the R.A. micrometer might be used on these in a similar manner. From (23), the condition for vertical transit is

$$2\sin\lambda_1\cos^2\beta_1+\sec(\lambda_1+\phi)\sin\phi = 0$$

or

$$\sin(2\lambda_1+\phi)+\sin\phi\tan^2\beta_1 = 0. \tag{33}$$

For the value of ϕ in question, λ_1 lies between 60° and 70°, as is also clear from Table I; this gives rather a large angle of incidence. From equation (16) we now have

$$\cos H\cos\delta = -2\sin^2\beta_1\sin\phi, \tag{34}$$

so that H is always a little greater than 6 hours, and the method is not much more rapid (if we use both tilts of the mirror-back) than the conventional one. However, it might perhaps still be valuable for comparison and check; it is clear from (31) that there would be some increase of accuracy as compared with a 4-hour interval, provided the large angle of incidence can be accepted.

Two important questions have so far been passed over in this discussion. The first is how the value of β_1 can be determined when it is, as it now will be, much too large for autocollimation in the tilted face to be possible. The second is how the azimuth of the mirror *in its cell* can be determined; this is, in effect, the same as asking what are the circle-readings when the tilted faces are vertical. It is these zeros from which λ_1 must in fact be reckoned, in applying equations (14), (23), (31), etc. If the azimuth of the mirror in its cell were zero, i.e. if the "ridge" of the roof were exactly perpendicular to the rotation axis, when one tilted face was vertical the other would be so also, though the front face would not necessarily be quite vertical as well, and though the two values of β_1 (for a roof-shaped back) might also be unequal. The values of β_1 could presumably be obtained by goniometer methods with sufficient accuracy; for example, the mirror might be rotated in its cell, round its own normal, through 90°, and the circle might then be read for autocollimation in the three faces. Alternatively, it may be obtained from R.A. measurements, using equation (15). The question of the zero of λ_1 is a little more difficult.

The effect of rotating the mirror through a small azimuth in its cell will be to increase the zero value from which λ_1 is measured, in the case of one half of the roof, and to decrease it for the other. If the mirror is rotated through an angle γ the change in λ_1 will be $\gamma\sin\beta_1$; there will also, in principle, be a change in the effective value of β_1 itself, but this is of the second order and can be neglected. From (14) we thus have, if $d_2\delta$ is the effect on δ of an azimuth error of the *mirror*,

$$\cos\delta\,.\,d_2\delta = 2\cos^2\beta_1\sin(2\lambda_1+\phi)d\lambda_1$$

or

$$d_2\delta = \sin 2\beta_1\sec\delta\,.\,\gamma\cos\beta_1\sin(2\lambda_1+\phi). \tag{35}$$

If, on the other hand, $d_1\delta$ is the effect on δ of an azimuth-error of the instrument, we have from (31) and (15)

$$d_1\delta = \sin 2\beta_1\sec\delta\,.\,a\cos\lambda_1\cos\phi. \tag{36}$$

The dependence on λ_1 is different in (35) and (36), so that if stars can in fact be observed over a wide range of λ_1 the two effects are separable. Indeed they are in principle separable even with a single star, provided it can be observed in both collimators. However, any such separation, whether using one collimator

and two stars, or two collimators and one star, will in general involve considerably tilted transits. It would require trial before one could say with certainty what accuracy could be obtained with large tilts, and we may therefore note that there are at least two other methods of attacking the same problem. On the one hand, we might employ any one of various known pieces of auxiliary apparatus to determine directly when a tilted face was vertical; and on the other, we might determine γ by comparison between the intrumental azimuths obtained by extra-meridianal declinations and the azimuths obtained in the standard manner from sub-pole right ascensions observed in the front face of the mirror.

It is clear that the mounting of the mirror must be such as to prevent any change in γ; however, if the method of pivot-testing already proposed is adopted, any change either in β_1 or γ will in fact be immediately observable.

Naturally, the pivot errors must be applied to these measurements. We may note in passing that pivot errors are sometimes tabulated in terms of their horizontal and vertical components, but only the component parallel to the altitude of the telescope is relevant in the case of a telescope-transit; the component perpendicular to this merely rotates the field (quite imperceptibly, of course). The same applies to a mirror-transit working in the meridian; the component perpendicular to the mirror-normal merely rotates the mirror, and is thus without effect even in principle. If, however, we are working off the meridian, both components matter; the one parallel to the normal (of the front face) effectively changes β_1, and the other one, when multiplied by $\sin\beta_1$, is directly a correction to λ_1, just as if the mirror actually rotated in its cell.

The whole question of determining fundamental azimuths by extra-merdianal observations of declination is of course a large one, and we shall not pursue it further at present, though we may note that, obviously, the extra-meridian formulae for level and for diurnal aberration would have to be employed. The proposal to use a mirror-transit is not really conditional upon the practicability of this particular application; it appears to have so many other advantages that it would be worth trying even if this one were found to be quite illusory. We will therefore return now to the main question, i.e. we shall confine ourselves from now on to the case when β_1 is very small.

The scale of the chronograph record will evidently vary in just the same way, with declination, as in the case of a telescope-transit. Thus the collimation correction, due to the fact that when the micrometer is in the position of the "standard contact" the line of collimation is not perpendicular to the rotation axis, is of the form $c\sec\delta$ just as before, c being the difference between the micrometer reading for the "standard contact" and the zero obtained as on p. 295 (both in seconds of arc). In addition, there is now the correction due to the fact that β_1 is not zero; from (15) we have

$$\begin{aligned} dH &= \cos\lambda_1 \sin 2\beta_1 \sec\delta \\ &\approx 2\beta_1 \cos\lambda_1 \sec\delta. \end{aligned} \tag{37}$$

Since now for small β_1, (14) becomes

$$\sin\delta \approx -\cos(2\lambda_1 + \phi)$$

we have

$$2\lambda_1 + \phi \approx \pi/2 + \delta \tag{38}$$

as is of course also geometrically obvious. Thus this correction is also expressible for any given β_1 purely as a function of δ, though it does not, of course, vary simply

as $\sec\delta$. Accordingly, it is only necessary to replace the usual correction table, which gives $c\sec\delta$, by one table for $c\sec\delta$ and a second one for $2\beta_1\cos\lambda_1\sec\delta$, and the complete collimation is allowed for essentially as before.

One undeniable drawback to the mirror-transit is that the telescope, and still more the observer himself, tends to obstruct the line of sight to any possible azimuth mark. It is true that one might in principle arrange a mark at 6° depression; if the observer could be so placed that not only was there still a clear view back past his body, but also no deleterious heating of the air by the proximity of his body, a depression of 6° could in fact be accepted; indeed, it would even perhaps be desirable in so far as the optical path could then be underground and so shielded from the heat of the Sun. However, unless the telescopes themselves are broken, a clear optical path would not be easy to arrange.

An alternative would be to provide a second pair of Y's on the main piers but about a foot higher than the mirror's pair, and somewhat further apart. If an axle (either rectangular or round) were laid in these, it could support a right-angle prism above the mirror in the manner sketched in Fig. 5, and azimuth marks could be observed by reflection in one or the other face of this prism.

It seems probable that the two sets of Y's would change their azimuth together,

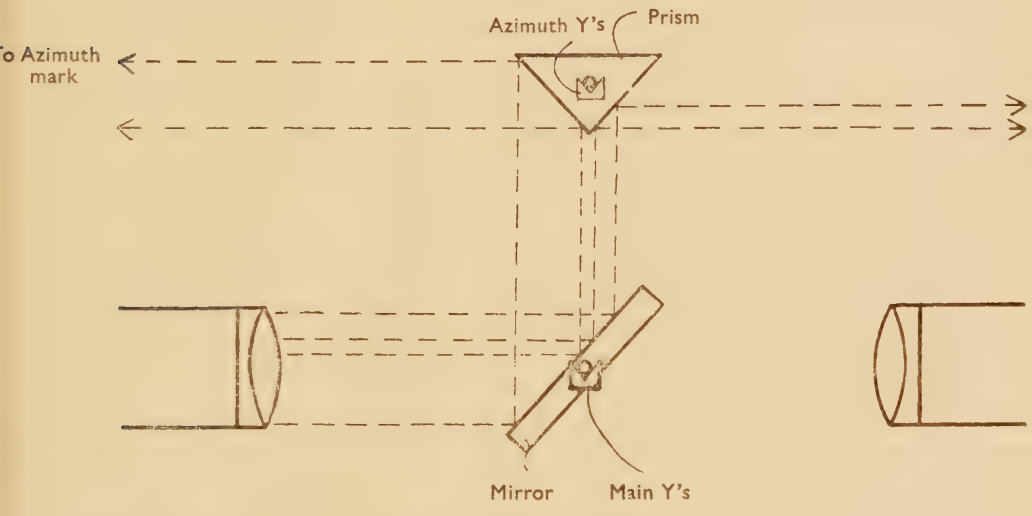


Fig. 5.

though admittedly if a change were due to one pier leaning over more north than the other it would have a larger effect on the azimuth Y's than on the main ones. Each collimator could observe, at the same time, one azimuth mark with the upper half of its aperture and the other with the lower; but it does not appear that this fact provides any additional check of value. The method is not altogether attractive; in particular, there is no obviously satisfactory place for this axle and prism when it is desired to observe near-zenith stars.

Of course, if it is in fact practicable to keep a running check on *fundamental* azimuths, by extra-meridianal observations, the need for azimuth marks largely disappears. They do, however, in favourable cases also provide a check on the cross-component of the polar movement, and they are therefore valuable even if there is no diurnal azimuth variation as such. It would clearly be preferable to have them if possible.

We are thus led to consider the use of broken telescopes. Clearly, if we can afford the necessary additional reflection, we can permanently "break" each of the telescopes, or break the beams even before we come to the telescopes; the line of sight to the azimuth mark, though it would still have to be depressed about 6°, would then be free from all interference by the observer's body. It is therefore worth noting that there is one other appreciable advantage in the use of broken telescopes.

If, namely, the telescopes are both broken towards the same side, say east, and both by slightly more than a right angle, it becomes possible to bring both observing positions close together. Modern methods of recording (for example on the 6″ transit circle at Washington) provide for all circle- and micrometer-readings to be taken by photography, so that one observer can do what previously needed two; but if the two collimator lenses are separated by 15 feet, and if the two focal lengths (as is perhaps not excessive) are now to be 30 feet, the observer will have to spend a good deal of his time moving between two observing positions 75 feet apart if the telescopes are not broken. It would clearly be better to have the positions close together, especially if the circle-settings were also made by a motor. One might, indeed, mount a long-focus pivot-telescope permanently between the other two, and take pivot-errors regularly after collimation, if desired; a few stars well observed are undoubtedly better than many poor ones. It should, however, be noted that there will be a tendency for the "main" mirror and the "breaking" mirror to act like a pair of crossed Nicols; if the mirrors were both unsilvered, there would in fact be serious extinction beyond the zenith. With silvered surfaces the effect is of course relatively much smaller, but the loss will still be larger than if polarization could be neglected.

In conclusion, we may note a few points in connection with the mounting of the mirror. We have seen that if it either rotates about its normal or tips laterally in its cell, the movement can be detected if we observe through the pivot-telescope; nevertheless, it is highly desirable to eliminate these movements if possible. In addition, it is absolutely essential to eliminate any rotation about an axis parallel to the main rotation axis of the instrument. It seems likely that the only satisfactory procedure would be to apply counterpoises so that the pressure of the mirror on each of the six ideally necessary restraining-points did not vary as the circle was rotated. If we are to be able to use both the front and the back of the mirror, only the marginal regions of these faces could be used for positioning; however, with a reasonably thick mirror this should not result in any appreciable flexure. The counterpoises might be attached to a belt running all round the edge, preferably fitted into a groove cut in the edge except where the small pivot-testing flats come; they should work with no slap or idle range at all, going over smoothly from pull to push as the sign of the load changes. If they are not to do any absolute positioning of the mirror themselves, it seems likely that they would have to be designed so that, in the absence of the six restraining-points, each one of them was in a neutral equilibrium rather than stable; but there does not seem to be any disadvantage in this. The necessary pressure on the six points would of course be provided by springs, not weights, and it should be possible to keep it quite small if the counterpoises are properly designed.

The six points themselves would probably be three on the front of the mirror near its edges, two on one of the side faces perpendicular to the axis (if the mirror

is rectangular) and one on one side face parallel to the axis. If in fact the mirror were round, it would be necessary to form special steps on its sides.

In any case, each of the six "points" should be so shaped (very nearly flat to the glass) that the area in contact, due to the elastic deformation of glass and "point", is as large as possible at the pressures used.

Royal Observatory,
Greenwich :
1946 *September* 4.

GALACTIC NOISE

R. v. d. R. Woolley

(Received 1947 February 11)

Summary

The electron velocity temperature of interstellar space is examined. It is found that while temperatures of order 12,000 deg. may be expected so close to early type stars that hydrogen is mainly ionized, in other regions, where metallic ions greatly outnumber hydrogen ions, the temperature is only of order 1000 deg. Galactic noise of intensity equal to that reported can only come from the hot regions, and probably only from regions where the hydrogen population considerably exceeds 1 per c.c. It is pointed out that if the observed noise is due to free-free transitions, the areas where it originates should exhibit bright hydrogen emission in observable quantities.

---

1. The existence of radiation with a wave-length of order one metre originating, apparently, in the galaxy has been known for some years.* Discussion of the subject has been recently renewed, and as great improvements have been made in radio technique in the last few years, it seems very likely that observations in these wave-lengths will yield useful information about the galaxy. The time is accordingly appropriate to review the theory of the emission of this radiation, which is related to the question of the temperature of interstellar space. Since the radiation in question is considered short-wave radiation by radio workers and is considered very long-wave radiation by spectroscopists, it seems undesirable to refer to it by a name embodying an idea of its wave-length. It has been called "noise" as it can actually be heard in earphones.

The temperature of interstellar space was discussed by Eddington in *The Internal Constitution of the Stars*. By temperature Eddington meant the parameter appearing in the Maxwellian velocity distribution of the free electrons. He supposed that this temperature was controlled by a balance between absorption of ultra-violet quanta and emission of radiation through "switches of electron orbits at encounters with atoms" or, as we should say now, free-free transitions. The free-free transitions have been supposed by Keenan and Henyey† to be the origin of galactic noise.

Eddington originally estimated the temperature of interstellar space at 10,000 deg., but later he suggested‡ that absorption in space of ultra-violet quanta would reduce the temperature.

Besides temperature, the degree of ionization in interstellar space is important. This has been investigated by Rosseland§ in general, and in detail having regard to the absorption of ultra-violet quanta by Strömgren.|| The last two writers calculate the ionization assuming an electron temperature. In the present

* Reber, *Ap. J.*, **91**, 621, 1940, who gives earlier references.
† *Ap. J.*, **91**, 625, 1940.
‡ *Observatory*, **60**, 99, 1937.
§ *Theoretical Astrophysics*, Chapter xxii, especially equation (139).
|| *Ap. J.*, **89**, 526, 1939.

paper simultaneous equations for the ionization and electron temperature are used.

2. We consider first the case where only one element (e.g. hydrogen) is present. Let there be n_e electrons, n^+ positive ions and n atoms per unit volume at a particular point in interstellar space, and let us suppose that there are $n_e M(v, T)dv$ electrons per unit volume with velocities between v and $v+dv$, where

$$M(v, T) = 4\pi(2\pi RT)^{-\frac{3}{2}} m^3 v^2 \exp(-mv^2/2RT).$$

We suppose that this distribution is brought about by elastic collisions. The distribution defines T, the temperature at the point concerned. Consider a point illuminated by light from a single object radiating like a black body at some temperature T^*. Then $J(\nu)$, the intensity of the radiation at the point, is given by

$$J(\nu) = \omega B(\nu, T^*) = \omega \cdot 2h\nu^3 c^{-2} (\exp h\nu/RT^* - 1)^{-1},$$

where ω is a number less than unity. (In practical applications, a very small number.)

Let there be $n_e n^+ M(v, T)\alpha(v)$ recombinations per unit volume per second between positive ions and electrons with velocities between v and $v+dv$, and let

$$\int_0^\infty M(v, T)\alpha(v)dv = \alpha(T).$$

By definition free-free transitions and elastic collisions do not affect the total number of free electrons per c.c., so that radiation ionizations (proportional to $n\omega$) are balanced by recombinations emitting radiation (proportional to $n_e n^+$). The equation which represents this balance is

$$\left(\frac{n_e n^+}{n}\right)\alpha(T) = \omega\left(\frac{n_e n^+}{n}\right)_{T^*}\alpha(T^*), \tag{2.1}$$

where $(n_e n^+/n)_{T^*}$ is the ionization in thermodynamical equilibrium at the temperature T^*, namely

$$2(2\pi mRT^*)^{\frac{3}{2}} \varpi h^{-3} \exp(-I/RT^*),$$

I being the ionization potential and ϖ the ratio of the statistical weights of the ionized and neutral states of the atom.* This is one relation involving the ionization $(n_e n^+/n)$ and the temperature T, ω and T^* being given, and we require a second relation.

Let the energy absorbed per c.c. per second by bound-free transitions be

$$4\pi n\omega \int_{\nu_0}^\infty k(\nu)B(\nu, T^*)d\nu, \tag{2.2}$$

where $h\nu_0 = I$, and let the emission be $4\pi n_e n^+ F(T)$. Since in full equilibrium the two must balance

$$\left(\frac{n_e n^+}{n}\right)F(T) = \int_{\nu_0}^\infty k(\nu)B(\nu, T)d\nu.$$

Again, let the energy absorbed from the radiation field by free-free transitions be

$$4\pi n_e n^+ \omega \int_0^\infty k'(\nu)B(\nu, T^*)d\nu$$

* *Cf.* Strömgren, *loc. cit.* equation (1), in which $\alpha(T) :: T^{-\frac{1}{2}}$.

and let the corresponding emission be $4\pi n_e n^+ F'(T)$.

Here $F'(T)=\int_0^\infty k'(\nu)B(\nu, T)d\nu$.

The energy balance is

$$n\omega\int_{\nu_0}^\infty k(\nu)B(\nu, T^*)d\nu+n_e n^+\omega\int_0^\infty k'(\nu)B(\nu, T^*)d\nu=n_e n^+\{F(T)+F'(T)\}. \quad (2.3)$$

Now when ω is very small $n_e n^+/n$ is of order ω and the term

$$n_e n^+/n\,.\,\omega\int_\nu^\infty k'(\nu)B(\nu, T^*)d\nu$$

is of order ω^2 and can be neglected. Neglecting this term in (2.3) and using (2.2) we get

$$\left(\frac{n_e n^+}{n}\right)\{F(T)+F'(T)\}=\omega\,.\left(\frac{n_e n^+}{n}\right)_{T^*}F(T^*) \quad (2.4)$$

the second equation which with (2.1) determines the ionization $(n_e n^+/n)$ and the temperature T, when ω and T^* are given.

We can at once eliminate the ionization and we have for the temperature

$$\{F(T)+F'(T)\}(\alpha(T))^{-1}=F(T^*)(\alpha(T^*))^{-1}. \quad (2.5)$$

Now

$$(n_e n^+/n)_T=2(2\pi mRT)^{\frac{3}{2}}h^{-3}\varpi\,\exp\,(-h\nu_0/RT)$$

and if we suppose that $k(\nu)=0$ for all $\nu<\nu_0$ and $k(\nu)=k(\nu_0/\nu)^n$ for all $\nu\geqslant\nu_0$ we find, neglecting $(RT/h\nu_0)^2$ and higher powers,

$$F(T)=CT^{\frac{1}{2}}\left(3-n+\frac{h\nu_0}{RT}\right), \quad \text{where } C=\frac{kh\nu_0^2R^{\frac{1}{2}}}{(2\pi m)^{\frac{3}{2}}c^2\varpi},$$

$$\alpha(T)=\text{const}\times\left(2-n+\frac{h\nu_0}{RT}\right).$$

Assuming for the moment that $F'(T)$ may be expressed $F'(T)=AT^{\frac{1}{2}}$ and that $A(RT/h\nu_0)$ can be neglected in comparison with C we get from (2.5)

$$\frac{T}{T^*}=\frac{C}{A+C}. \quad (2.6)$$

3. Kramers's formula gives the emission per c.c. due to free-free transitions as

$$n_e n^+\frac{2^7\pi^3}{(6\pi)^{\frac{3}{2}}}\,.\,\frac{e^6}{hc^3m}\left(\frac{RT}{m}\right)^{\frac{1}{2}},$$

which is of form $F'(T)=AT^{\frac{1}{2}}$.

Since for hydrogen $\varpi=\frac{1}{2}$, and for the Lyman continuum

$$k\nu_0^3=\frac{2^6\pi^4}{3\sqrt{3}}\,.\,\frac{me^{10}}{ch^6} \quad \text{and} \quad \nu_0=2\pi^2me^4h^{-3}$$

for hydrogen $A=\frac{1}{2}C$ and for interstellar material consisting solely of hydrogen illuminated by a single black body radiating at a temperature T^*

$$T=\tfrac{2}{3}T^* \quad (3.1)$$

by (2.6). We have supposed that the higher continua (Balmer, Paschen, etc.) do not influence the result, and we have supposed that we are dealing with a

region where the Lyman continuum has not been removed by absorption in the interstellar gas.

In the case of a pure hydrogen interstellar gas illuminated by a number of stars each radiating like a black body, (2.4) is replaced by

$$\left(\frac{n_e n^+}{n}\right)\alpha(T)=\Sigma_p\omega_p\left(\frac{n_e n^+}{n}\right)_{T_p^*}\alpha(T_p^*) \tag{3.2}$$

while (2.5) can be reduced to

$$\left(\frac{n_e n^+}{n}\right)\alpha(T)\ .\ T=\frac{C}{A+C}\ .\ \Sigma_p\omega_p\alpha(T_p^*)T_p^*\left(\frac{n_e n^+}{n}\right)_{T_p^*} \tag{3.3}$$

having regard to (3.2). The stars are weighted according to the ionization they produce, and stars of type Bo and earlier are most important. Equations (3.2) and (3.3) together give the ionization $n_e n^+/n$ and the temperature T when each ω_p and T_p^* are given.

Eddington's original estimate of the radiation field was: 5 per cent due to stars at 18,000 deg.; 10 per cent, 12,000 deg.; 20 per cent, 9,000 deg.; 40 per cent, 6,000 deg.; and 25 per cent, 3,000 deg. This is equivalent to

$$J(\nu)=\omega'\{5B(\nu,\ 18{,}000\text{ deg.})+10B(\nu,\ 12{,}000\text{ deg.})+\text{etc.}\} \tag{3.4}$$

and it is easily verified that the terms due to stars at temperatures other than 18,000 deg. are very small in (3.2) and (3.3). The result of adopting Eddington's composition is very nearly $T=\frac{2}{3}\times 18{,}000$ deg. The ionization may be found by calculating the value of ω' in (3.4) from Eddington's estimate of the energy density, namely, $7{\cdot}67\times 10^{-13}$ ergs per c.c., which gives $\omega'=1{\cdot}1\times 10^{-16}$.

A very direct way of determining ω for a particular star may be mentioned. For the Sun as seen from the Earth $\omega=5{\cdot}41\times 10^{-6}$ so that for a star having the same radius as the Sun and distant 10 parsecs $\omega=1{\cdot}35\times 10^{-17}$. If in addition the star has the same effective temperature as the Sun its apparent bolometric magnitude would be $4^{\text{m}}{\cdot}85$. Hence, since for any star ω is proportional to r^2p^2 (where p is the parallax) while the apparent luminosity is proportional to $r^2p^2T_e^4$, for a star of apparent bolometric magnitude m

$$\begin{aligned}\log\omega &= -17{\cdot}13+0{\cdot}4(4{\cdot}85-m)+4\log T_\odot-4\log T_e\\ &= -0{\cdot}15-0{\cdot}4m-4\log T_e.\end{aligned}$$

For a B type star with $T_e=18{,}000$ deg. and $m=1{\cdot}0$, $\log\omega=-17{\cdot}57$. An actual count in Schlesinger's *Catalogue of Bright Stars* shows that the O stars and stars of type Bo to B5 inclusive are equivalent to 31 stars of visual magnitude 1·0. (The catalogue is complete only to the sixth magnitude but as the maximum contribution to total light comes near $4^{\text{m}}{\cdot}0$ in these spectral types, it is safe to ignore the contribution by faint stars omitted from the catalogue.) Assuming that the mean correction to bolometric magnitude is $1^{\text{m}}{\cdot}0$ and that all the stars may be taken together as if each had an effective temperature of 18,000 deg., we get $\omega=2{\cdot}10^{-16}$ in good agreement (considering the rough nature of the estimate, necessarily so on account of the uncertainty in T_e) with Eddington's $\omega=5{\cdot}5\times 10^{-16}$. With $\omega=5{\cdot}5\times 10^{-16}$ and $T=12{,}000$ deg. the ionization of hydrogen is, by (3.2), $\frac{n_e n^+}{n}=525$, so that even with a higher value of n_e such as $n_e=10$ the hydrogen is mainly ionized.

If there are two elements present, equation (2.1) is replaced by

$$\left(\frac{n_e n_1^+}{n_1}\right)\alpha(T)=\omega\left(\frac{n_e n_1^+}{n_1}\right)_{T^*}\alpha(T^*), \quad \left(\frac{n_e n_2^+}{n_2}\right)\alpha(T)=\omega\left(\frac{n_e n_2^+}{n_2}\right)_{T^*}\alpha(T^*) \quad (3.5)$$

and (2.3) replaced by

$$n_e n_1^+\{F_1(T)+F_1'(T)\}+n_e n_2^+\{F_2(T)+F_2'(T)\}$$
$$=\omega n_1\int_{\nu_1}^{\infty}k_1(\nu)B(\nu,\ T^*)d\nu+\omega n_2\int_{\nu_2}^{\infty}k_2(\nu)B(\nu,\ T^*)d\nu, \quad (3.6)$$

which, using (3.5), can be expressed

$$(\alpha(T))^{-1}\{n_1^+(F_1(T)+F_1'(T))+n_2^+(F_2(T)+F_2(T))\}$$
$$=(\alpha(T^*))^{-1}\{n_1^+F_1(T^*)+n_2^+F_2(T^*)\}. \quad (3.7)$$

Hence in a mixture of hydrogen and metals in which hydrogen supplies most of the *ions*, or $n_1^+ \gg n_2^+$, the metals may be neglected so far as the temperature is concerned. This would be the case in a typical region of interstellar space if the B stars radiated like black bodies at 18,000 deg. in the far ultra-violet and if there was no appreciable absorption of the Lyman continua between the stars and the point considered. But as Eddington himself pointed out* the first assumption is open to considerable doubt, and the hydrogen in interstellar space itself will absorb the Lyman continua (and convert the quanta into lower energy quanta by fluorescence).

In a region in interstellar space where all the quanta in the Lyman continuum are absent, the effective value of ω for hydrogen is zero, n_H^+ is zero and (3.7) reduces to

$$\frac{T}{T^*}=\frac{C_m}{A+C_m},$$

where C_m is a suitable mean for the metals. The emission coefficient A is unaltered (if the metals are, on the whole, singly ionized) but C_m is proportional to $k\nu_0^2$ so that

$$A/C_m=\tfrac{1}{2}\,.\,(k\nu_0^2)_H/(k\nu_0^2)_{\text{metals}}\times\varpi/\tfrac{1}{2}.$$

It is impossible to calculate C_m with certainty since so little is known about absorption coefficients. The only metals whose absorption coefficients have been directly measured are *Na*, *K*, *Rb*, *Cs*, and the value is less than one-tenth of the value for hydrogen.† But according to Bates and Massey‡ the low coefficients of these alkali metals are abnormal, the abnormality not being shared by *Ca* (but Ca^+ exhibits it.) Even theoretical calculations do not extend at present to the heavier elements with complicated spectra. Again, the composition of interstellar space can only be guessed. According to Swings§ the only metals which have been identified with certainty are Ca^+, *Ca*, *Na*, *K*, Ti^+. From its stellar abundance one would expect *Mg* to be present, almost entirely in the form Mg^+, of which the ultimate lines are inaccessible (near 2800 A.). The absorption of Mg^+ is presumably "abnormal" like Ca^+ but as the ionization potential is greater than that of *H* the ion is inert in the present connection when the Lyman continuum is absent.

* *Observatory*, **60**, 99, 1937.
† *Cf.* Page, *M.N.*, **99**, 385, 1939.
‡ *Proc. R.S.*, **177**, 329, 1941.
§ *Annales d'Astrophys.*, **1**, 42, 1938.

We may make a compromise between "abnormal" and "normal" absorption and suppose that they are roughly equally represented, and for the sake of a round figure take $k = 0{\cdot}4k_H$. Then with $\nu = \frac{1}{2}(\nu_0)_H$ and $\varpi = 1$ we get $A/C_m = 1/10$. With no Lyman continuum the B stars are not weighted so heavily in comparison with later types. With $T^* = 11{,}000$ deg. we get $T = 1{,}000$ deg. This figure is very uncertain on account of the uncertainty in C_m.

Strömgren* has calculated the ionization of hydrogen at various distances from stars of different spectral types assuming certain electron temperatures. He finds in each case a critical distance at which the hydrogen ionization falls off sharply. If we assume a large abundance of hydrogen in interstellar space (similar to the abundance found in the Sun and the stars) we should expect, on the argument in this paragraph, an abrupt change in the electron temperature from $\frac{2}{3}T^*$ inside the region of high hydrogen ionization to about 1,000 deg. outside it.

4. In the noise wave-lengths stimulated emission is much greater than unstimulated emission, so that care must be paid to the distinction between the two types of emission. By supposing that the probabilities of transitions between velocities v, v' are related by Einstein coefficients in the same way as bound-bound transitions and then integrating over suitable v, v' pairs to give ν radiation, it can be shown that if the absorption per c.c. from $I(\nu)$ is

$$n_e n^+ A_{12}(\nu, T) I(\nu) \tag{4.1}$$

then the emission *in the same direction as* $I(\nu)$ is

$$n_e n^+ A_{12}(\nu, T) \exp(-h\nu/RT) \cdot \left\{\frac{8\pi h\nu^3}{c^2} + I(\nu)\right\}. \tag{4.2}$$

The differential equation for the transfer of radiation is

$$\frac{dI(\nu)}{k'(\nu)ds} = -I(\nu) + B(\nu, T) \tag{4.3}$$

if

$$k'(\nu) = A_{12}(\nu, T) \cdot (1 - e^{-h\nu/RT}). \tag{4.4}$$

Now, the unstimulated emission per c.c. per sec. in the interval ν to $\nu + d\nu$ is

$$j_\nu d\nu = n_e n^+ \frac{2^7\pi^3}{(6\pi)^{\frac{3}{2}}} \cdot \left(\frac{m}{RT}\right)^{\frac{1}{2}} \frac{e^6}{c^3 m^2} \exp\left(-\frac{h\nu}{RT}\right) \times g(\nu) d\nu, \tag{4.5}$$

where $g(\nu)$ is the "Gaunt factor" which is nearly equal to unity in the visible spectrum but attains large values in very long wave-lengths. In fact, as $\nu \to 0$, $g(\nu) \to \frac{3}{\sqrt{\pi}} \ln(E/h\nu)$, where E is the average free energy of the electrons. (In deriving the constant A from $AT^{\frac{1}{2}} = \int_0^\infty j_\nu \, d\nu$ we have taken $g(\nu) = 1$).

Now

$$j_\nu = n_e n^+ \cdot \frac{8\pi h\nu^3}{c^2} A_{12}(\nu, T) \cdot \exp(-h\nu/RT)$$

(since it is the unstimulated term in (4.2)), so that $A_{12}(\nu, T)$ can be found from (4.5) and then $k'(\nu)$ can be calculated from (4.4).

* *Loc. cit.*

We have, in fact, for the absorption coefficient *per centimetre* when $h\nu/RT$ is small,

$$k'(\nu)=n_e n^+ \frac{2^4\pi^3 e^6}{c(6\pi mRT)^{\frac{3}{2}}} \cdot \nu^{-2} g(\nu). \tag{4.6}$$

Let $\int_0^x k'(\nu)dx=\tau_\nu$. Looking in any direction in space τ_ν will tend to a limit t_ν when we reach the end of the galaxy. In some noise wave-lengths t_ν may be small. From (4.3) the radiation received at the Earth per sq. cm. per sec. from unit solid angle (from isothermal interstellar space) is

$$S_\nu = B(\nu, T) \cdot (1-e^{-t_\nu}). \tag{4.7}$$

Now for $\nu=10^8$ (100 megacycles) and $T=12{,}000$ deg., $g(\nu)\sim 8$ and $k\sim 2\cdot 10^{-2}$ per kiloparsec, so that as the diameter of the galaxy is of order 50 kiloparsecs, the optical depth is of order unity if $n_e=n^+=1$ and the galaxy is all "hot". For cold space with $T=1000$ deg., $k\sim n_e n^+$ per kiloparsec but (if the composition of interstellar space is the accepted stellar mixture) there is one metal atom to about 40 high ionization atoms (that is, *H*, together with *O*, *He*, etc.), so that if only the metals are ionized and the hydrogen population is 1 per c.c., then $n_e=n^+=1/40$ and $k\sim 10^{-3}$ per kiloparsec—cold interstellar matter having accordingly one-twentieth of the opacity to noise of hot interstellar matter. The noise emissivity of the hot matter is, on the other hand, two hundred times that of the cold matter.

Now the observed maximum noise at 64 megacycles per sec., according to Hey, Phillips and Parsons is about 10^{-17} watts per square metre per kilocycle band width per unit solid angle, which is the same order of magnitude as black body radiation at 10,000 deg. Hence, in the (limited) directions from which the full noise comes, we must be looking at unit optical depth or greater by (4.7). It seems likely that in these regions n_e and n^+ must considerably exceed unity. For example, if $n_e=n^+=10^2$ unit optical depth is attained in less than a tenth of a kiloparsec, which is reasonably small. In planetary nebulae the hydrogen population is estimated at 10^4 per c.c.

To sum up the discussion in this paragraph, if noise is due to free-free transitions, it must certainly come from hot regions in interstellar space, and almost certainly from regions where the density is above average.

5. If the noise is due to free-free transitions, and if "cold" regions (regions of low hydrogen ionization) do not contribute appreciably to the noise, *an observed amount of galactic noise from a particular part of the sky implies a minimum bright hydrogen emission from that part of the sky*. If this is not revealed by observation, it will be necessary to conclude that the galactic noise is not due to free-free transitions. Perhaps we should be prepared for this, as the conspicuous variations in solar noise which have been reported may imply that a good deal of the noise is electrical rather than thermal in character. Hey* has reported variations in *galactic* noise.

Returning to the view that noise *is* due to free-free transitions, the emission is given by (4.5). Now from (4.7) if t_ν is small

$$S_\nu = B(\nu, T) \times t_\nu = \frac{j_\nu}{n_e n^+} \int n_e n^+ dx = \frac{j_\nu}{n_e n^+} \times \sigma, \text{ say,} \tag{5.1}$$

* *Nature*, **158**, 234, 1946.

and the value of σ deduced from an observed S by (5.1) is a minimum. According to Cillié * the number of electron captures in the third hydrogen level per c.c. per sec. is

$$N_3 = n_e n^+ \frac{2^9\pi^5}{(6\pi)^{\frac{3}{2}}} \cdot \left(\frac{m}{RT}\right)^{\frac{3}{2}} \cdot \frac{e^{10}}{27m^2c^3h^3} \exp(\chi_3/RT) \cdot Ei(\chi_3/RT), \qquad (5.2)$$

where χ_3 is the energy necessary to ionize the third state, namely, 1·5 volts.

Cillié supposes that each capture in the third state ultimately releases an $H\alpha$ quantum, because if an $L\beta$ quantum is emitted it is reabsorbed in interstellar space. This is the minimum number of $H\alpha$ quanta emitted, as some will come from captures on higher levels.

The minimum number of $H\alpha$ quanta received per sq. cm. per sec. per unit solid angle of sky observed is therefore

$$Q = \frac{1}{4\pi} \cdot \frac{N_3}{n_e n^+} \cdot \sigma$$

and using (4.1), (4.3) and (5.1) with $e^{h\nu/RT} \sim 1$

$$Q = \frac{\pi S_\nu e^4}{27h^3} \cdot \left(\frac{m}{RT}\right) \cdot \exp(\chi_3/RT) Ei(\chi_3/RT) \cdot 1/g(\nu). \qquad (5.3)$$

The quantity $T^{-1} \exp(\chi_3/RT) Ei(\chi_3/RT)$ is not sensitive to the choice of T. For $T = 12{,}000$ deg., $\exp(\chi_3/RT) Ei(\chi_3/RT) = 0{\cdot}540$, $g(\nu) \sim 8$ and then $Q \sim 8{\cdot}10^{23}\ S_\nu$ quanta per sq. cm. per sec.

Now according to Reber † the energy actually received is of order 10^{-17} ergs per sq. cm. per sec. per square degree per kilocycle band width, or $S_\nu \sim 3{\cdot}10^{-17}$ giving $Q \sim 2{\cdot}10^7$ quanta per sq. cm. per sec. per unit solid angle. Hey, Phillips and Parsons ‡, using a longer wave-length, find $3{\cdot}10^{-18}$ in Reber's units for the maximum noise intensity (in Sagittarius).

The predicted value of Q is within reach of observation. Struve § has described a very fast spectrograph with which he and his co-workers have detected extremely faint hydrogen emission from various regions in the sky. The faintest nebula so far observed || is estimated to have one atom in the third hydrogen state per square centimetre of surface presented to the Earth, the corresponding value of Q being $3{\cdot}10^6$ quanta per sec. per sq. cm. per unit solid angle subtended. While Struve thinks that this value of Q is subject to uncertainty, there is enough margin between the two values of Q for us to assert that a noise level of 10^{-17} ergs as specified above ¶ should certainly be accompanied by observable $H\alpha$ emission if the noise is due to free-free transitions.

The author acknowledges with pleasure his indebtedness to Dr D. F. Martyn for calling attention to the problems presented by galactic noise and for very valuable discussions on the proper treatment of stimulated emission.

Commonwealth Observatory,
Canberra, Australia :
1947 *January* 17.

* *M.N.*, **92**, 820, 1932.
† *Ap. J.*, **91**, 621, 1940.
‡ *Nature*, **157**, 296, 1946.
§ *Ap. J.*, **87**, 559, 1938.
|| *A. J.*, **89**, 517, 1939.

¶ As these observations are made with instruments of *very low* resolving power, the maximum values from small areas will be much greater than the observed values if the noise has any structure, i.e. comes from areas of order 1 sq. degree or less.

A SPECTROPHOTOMETRIC STUDY OF MAXIMA IN THE EMISSION LINES OF NOVA HERCULIS 1934

M. Khairy Aly

(Received 1947 March 8)

Summary

Sixteen emission lines in the red and green regions of the spectrum of Nova Herculis 1934 have been photometrically studied during the period from the beginning of 1934 January till late in 1935 April. Comparison of the two components of these lines have been made whenever possible, and variations of their relative intensities have been correlated with light variations of the nova. An anomaly has been found in the relation between the intensity ratio R/V and magnitude in the case of hydrogen lines. Most of the intensity measures of emission lines, in particular those of oxygen, calcium and some iron lines have been found to agree with work of previous investigators. Among the measurements those of the [*OI*] lines are the most reliable. Periodic variations in the light of the nova were correlated to the periodic variations in the relative intensities of the line components. It is suggested that the general behaviour of the emission components can be explained in terms of hot jets of matter of different composition streaming in opposite directions outwards from the star and forming stratified layers. The condensations resulting from emitted matter appear to have been present at an early date after the nova outburst as shown by the presence of maxima, the same is probably true of the condensations observed visually at later dates.

---

1. *Introduction.*—Most of the emission lines in Nova Herculis 1934 show double maxima for quite a long period in the different stages of the brightness of the nova. The object of this paper is to make a quantitative study of the variations in the relative strengths of the two emission components at different stages of the nova with respect to each other and with respect to the continuous background; also to try to picture the nova's development in the light of the general behaviour of the lines. A study of the contours and intensities of some emission and absorption bands was made by E. G. Williams *, especially for some hydrogen and calcium lines. Later N. T. Bobrovnikoff † gave a detailed investigation of the structure of the bands in the visual region. In this nova some of the emission bands showed more than two peaks on some dates due to the interference by displaced absorption lines of various origins. However, it was nearly always possible to identify the two relevant maxima by the aid of radial velocity measures which Professor Stratton kindly put at my disposal. As is the case in all novae, absorption bands are present side by side with emission bands making it sometimes difficult to determine accurately the level of the continuous background.

2. *Material.*—77 plates were used in this work; 19 of which were taken at Mount Wilson with the Cassegrain one-prism spectrograph and the small spectrograph, both attached to the 100-inch telescope, and with the three-prism ultra-violet spectrograph on either the 60 or 100-inch telescopes; 58 plates were taken at the Solar Physics Observatory, Cambridge, with the two-prism

* *M.N.*, **99**, 91, 1938.

† *Contr. Perkins Obs. Ohio*, No. 2, 1935 August.

spectrograph on the 25-inch Newall telescope for the photographic region and with the Littrow spectrograph on the 15-inch Huggins refractor for the visual region.

3. *Reduction of the plates.*—Mount Wilson plates had been standardized with a tube densitometer; in some cases they provided on the same plate more than one spectrum of the nova with different exposure times. In obtaining intensities for the very strong lines, in particular for $H\alpha$ and certain forbidden oxygen lines ([*OI*] $\lambda\lambda$ 6364, 6300), use was made of the different exposures on the same plate. From a slightly exposed spectrum the ratio of intensity between $H\alpha$ and either of the less dense forbidden lines of oxygen was obtained first; another more heavily exposed spectrum of the same plate was used then to derive the intensity of either of the oxygen lines relative to the continuous background. Finally the intensity of $H\alpha$ was deduced with respect to that of the continuous spectrum. In the comparison of $H\alpha$ with the oxygen lines allowance was made for the variation in the intensity of the continuous spectrum between λ6360 and λ6550 by assuming black-body radiation for the nova. Three values for the colour temperature of the nova, namely 10,000 deg. K., 11,500 deg. K. and 14,000 deg. K., were taken, corresponding to the photographic magnitudes 3 m., 3·5 m. and 4 m.* which the nova attained during the period with which we are concerned. The Cambridge plates had been standardized by means of a spectrograph employing a wedge. Star and image spectra were taken on two halves of plates which were developed together.

Both Mount Wilson and Cambridge plates were run through the microphotometer of the Solar Physics Observatory and intensities of the spectral lines were deduced from the photometer deflections, using measurements of the standard densities on the plates in the usual manner.

4. *The continuous spectrum.*—The measurement of the intensity of the continuous spectrum was in some cases one of the chief difficulties of the present work. The difficulty arose from the continuous spectrum being heavily affected by strong and wide emission and absorption bands, also from considerable variations in intensity, especially in the $H\alpha$ region of the spectrum. These variations have been measured by N. T. Bobrovnikoff †, who defined the position of maximum intensities of the continuous spectrum, especially in the red part of the spectrum and its effects on colour temperature determinations. However, in this investigation these variations will have no serious effect on the measures of the relative intensities of the components of any one of the emission bands, though they may affect the derived values of the absolute intensities of $H\alpha$.

To give an idea of the effect of choosing different temperatures of the nova on the values of central intensities for the two components of $H\alpha$, some results are presented in Table I in which intensities are expressed in units of the continuous background.

5. *Results of photometric reduction.*—Table II shows a list of all permitted and forbidden emission lines which have been studied here grouped in multiplets with their origins, lower and upper excitation potentials and their multiplets.

Most of these lines are possibly affected by blending with other displaced absorption or emission lines for some part of the period under investigation.

* F. L. Whipple and C. P. Gaposchkin, *Harvard Astronomical Circular*, No. 412, 1937: later referred to in text as *H.A.C.*

† *Op. cit.*

TABLE I

Intensities of Hα components corresponding to three different colour temperatures of the nova

Plate No.	10,000 deg. K.		11,500 deg. K.		14,000 deg. K.	
	Red	Violet	Red	Violet	Red	Violet
1212	18·9	24·8	19·7	26·0	22·8	30·4
1213	14·0	12·4	14·6	13·1	16·9	15·2
1255	10·1	8·1	10·6	8·4	12·2	9·8
1256	6·1	5·1	6·4	5·3	7·4	6·2
1259	10·8	13·4	11·3	14·0	13·1	16·2

TABLE II

List of emission lines

Wave-length A.	Origin	Excitation Potential in Volts		Multiplet
		Lower	Higher	
6562·82	*H*	10·15	12·04	2^2P^0–3^2D
4861·33	*H*	10·15	12·69	2^2P^0–4^2D
6363·88	[*O*I]	0·02	1·96	$2p^3P$–$2p^1D$
6300·23	[*O*I]	0·00	1·96	
5577·35	[*O*I]	1·96	4·17	$2p^1D$–$2p^2S$
5754·8	[*N*II]	1·89	4·04	$2p^1D$–$2p^1S$
5534·86	*Fe*II	3·23	5·46	b^2H–z^4F^0
5316·61	*Fe*II	3·14	5·46	a^4G–z^4F^0
5275·99	*Fe*II	3·19	5·52	
5234·62	*Fe*II	3·21	5·57	
5197·57	*Fe*II	3·22	5·57	
5169·03	*Fe*II	2·88	5·27	a^6S–z^6P^0
5018·43	*Fe*II	2·88	5·34	
4923·92	*Fe*II	2·88	5·39	
4233·17	*Fe*II	2·57	5·49	b^4P–z^4D^0
3933·66	*Ca*II	0·00	3·14	4^2S–4^2P

The blending might affect either of the red or violet components or both of them. Such possible effects of blending will now be indicated in the case of some of the lines.

Hβ: After the middle of 1935 March the red maximum may be stengthened by the violet maximum of [*Fe*II] λ4874·5. The two lines will be about 3 A. apart, however, and other hydrogen lines, particularly *Hδ*, exhibited the same emergence of a red maximum.

[*N*II] λ5755: This line may be affected during 1935 March by absorption VI of *N*II λ5767·4 which would weaken the red maximum of λ5755 and occasionally by absorption VIII of *N*II λ5767·4 which would weaken the violet maximum of λ5755. On the other hand the emission of *N*II λ5767·4 might make the red maximum of λ5755 too strong and the bright emission band too wide.

[*OI*] λ6364: The violet maximum at λ6358·1 might be weakened by absorption IV of *Si*II λ6371·36 up to the end of February and probably some time in 1935 January. The absorption of *Si*II λ6347·1, which goes nearly parallel with the absorption of *Si*II λ6371·36, could be seen in Stratton's Atlas till the end of February.

*Fe*II λ5535: This line may be affected during March by the *N*II emission lines λλ5530·3, 5535·4, 5543·5. The violet maximum of λ5535 might also be weakened by absorption VI of *N*II λ5535·4.

*Fe*II λ5316·6: This line might have its violet maximum weakened occasionally during March by absorption VIII of *N*II λ5321·0.

*Fe*II λ5276·0: It might have its violet maximum blended in the latter half of March with that of [*Fe*II] λ5273·4, and in January its red maximum may fall together with the violet maximum of *Fe*II λ5283.

*Fe*II λ5169: This line must be watched carefully for correct identification after the middle of March owing to the growing strength of the violet maximum of [*Fe*II] λ5158. There is also an extra pair of maxima due to an unidentified source about 5173 A.

*Fe*II λλ4923·9, 5018·4: These lines might be strengthened, relative to the continuous spectrum during March by the bright band of *He*I at λλ4921·9, 5016·7, these being emissions VI and VIII.

*Fe*II λ4233: It may have its red maximum at λ4237·8 weakened in January by absorption II of *Cr*II λ4242·4, and strengthened or masked in the latter half of March by the violet maximum of [*Fe*II] λ4244·4.

In Fig. 1 the actual values of the central intensities of both red (in full lines) and violet (in broken lines) components are plotted against time; a photographic light curve * is given at the top to show the correlation of intensities with light variations of the nova. The hydrogen and forbidden oxygen lines, together with the means of the two iron multiplets a^4G–z^4F^0 and a^6S–z^6P^0 and some other individual lines, are shown.

In Fig. 2 the intensity ratios R/V for individual observations are plotted in the same way.

To represent more clearly the difference in behaviour between the different lines and multiplets, Fig. 3*a*, ... 3*f* give the individual observations of R/V plotted against magnitude.

The same type of relationship appears to exist between values of R/V and magnitude for $H\alpha$ and $H\beta$. The straight lines drawn in Fig. 3*a*, 3*b* have the same slope relative to the axes. The two points at the right bottom corner of the $H\alpha$ diagram correspond to decreasing brightness of the nova at the deep minimum when obviously some new factor is operating, however, and one of them is not very reliable. They show roughly a reversal in inclination. Thus the hydrogen lines show an increase in intensity of their red components relative to the violet ones with decreasing brightness of the nova till this reached the fifth magnitude.

The same effect has been observed in the diagram showing the behaviour of the ionized iron lines belonging to the multiplet a^6S–z^6P^0. The values of R/V appear to show first a growing intensity of the red component with respect to the violet with decreasing brightness of the star, indicated by the broken lines in the diagram, until the magnitude of the nova is 3·5, when the reverse occurs.

* *Publ. L'Obs. Lyon*, **1**, 16, 1935.

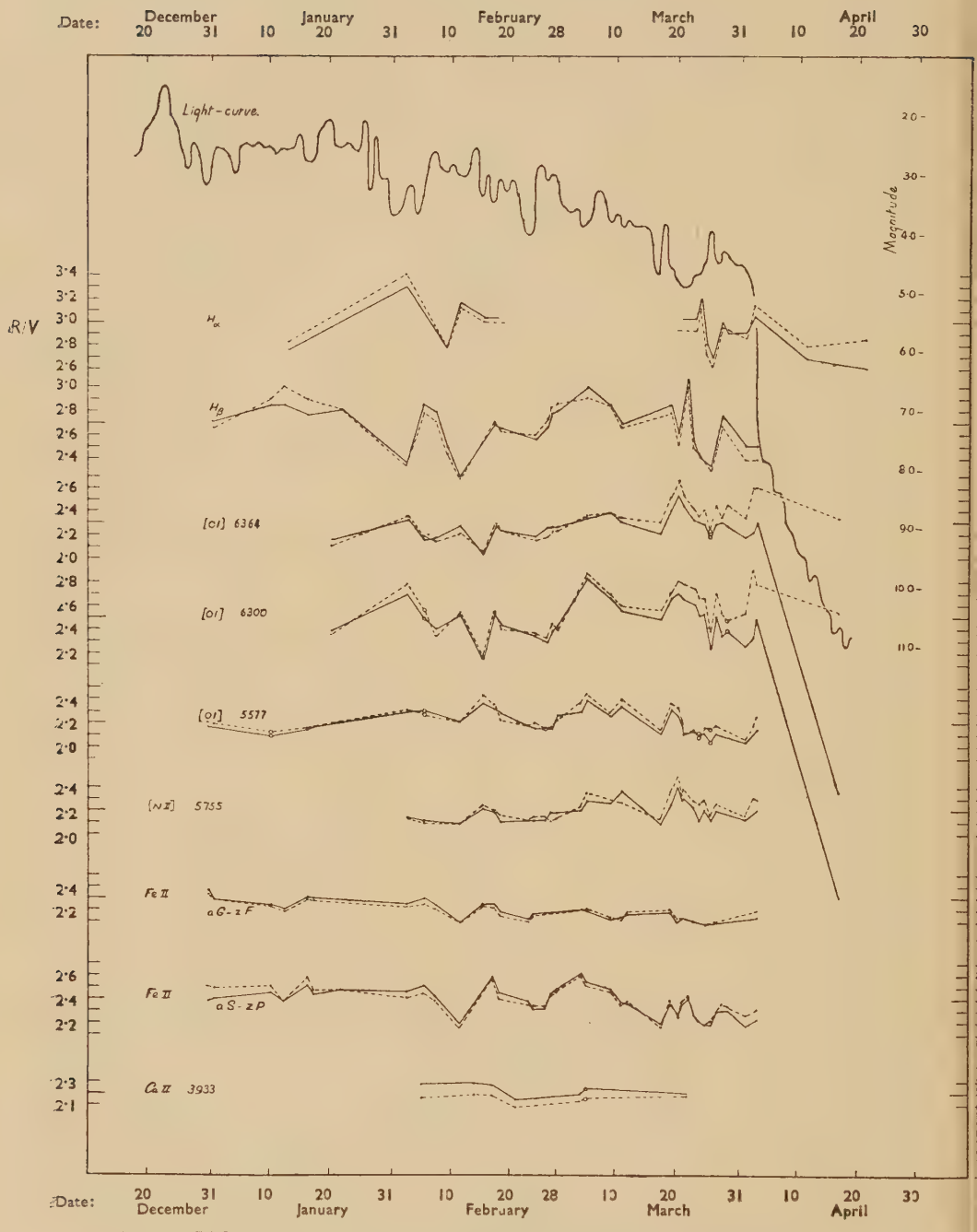


FIG. 1.—*Light curve and intensities of red (full) and violet (broken) components.*
o—*Low weight.*

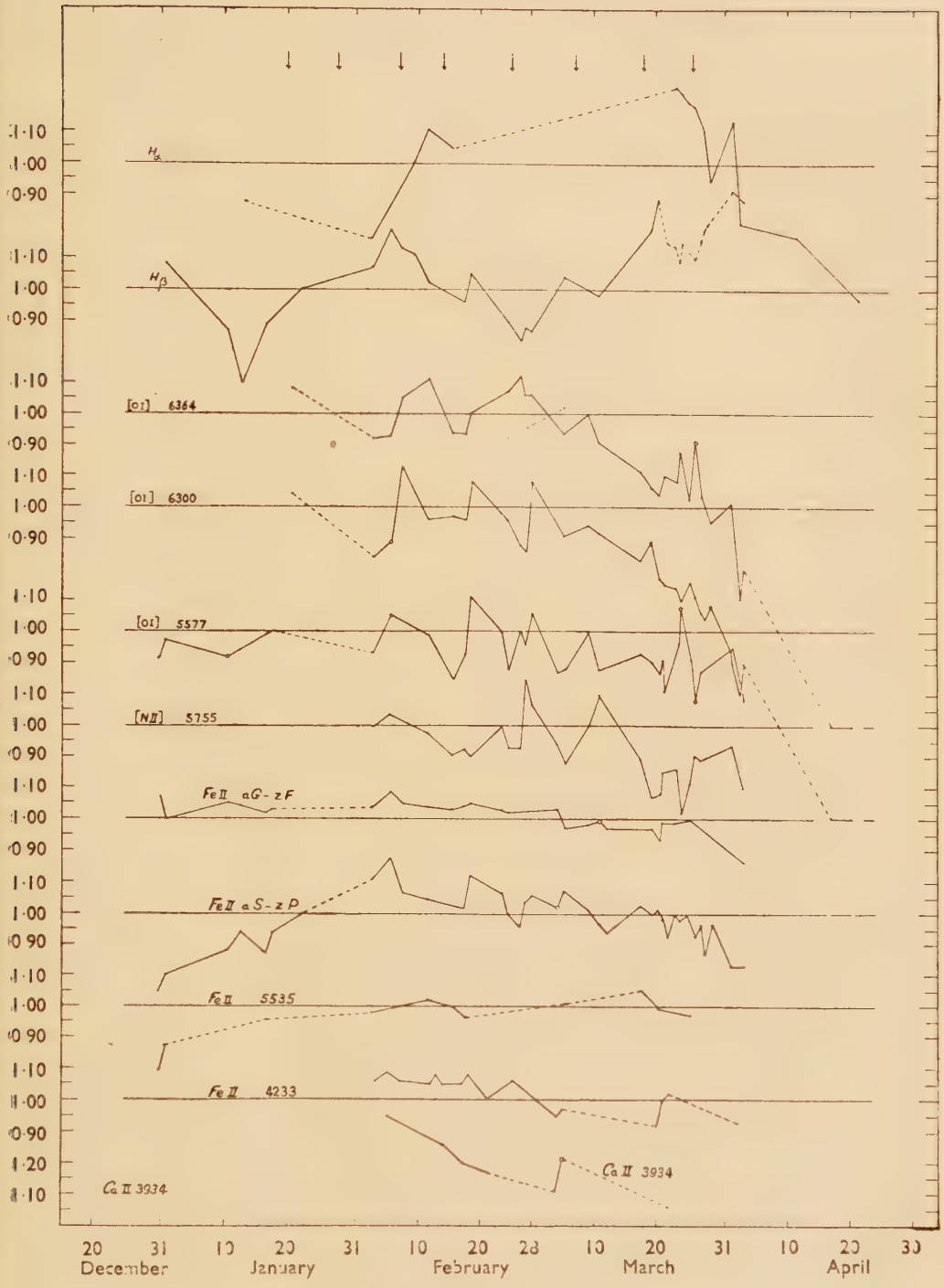


FIG. 2.—*Ordinates : R/V.*
Abscissae : Date.
Arrows at head of diagram indicate dates of maxima in the light curve.
o —*Low weight.*

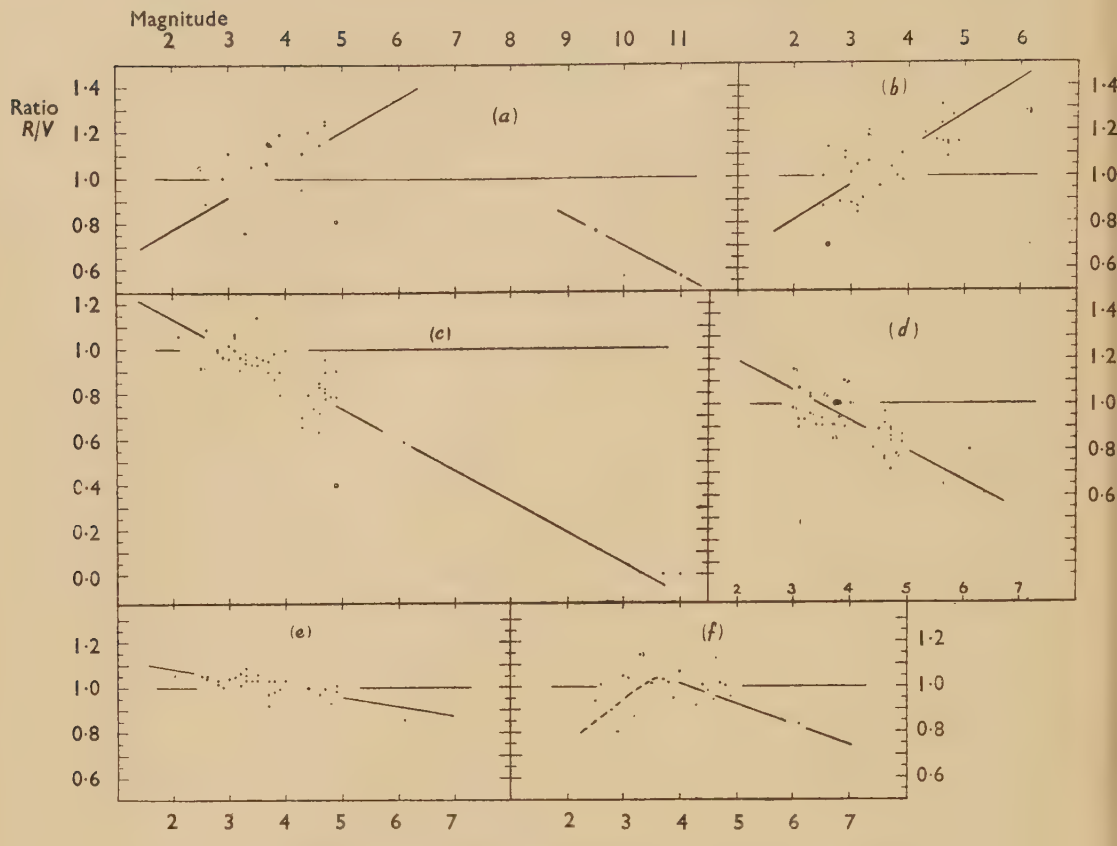


FIG. 3.—*Ratio R/V (ordinates) and Magnitude (abscissae) of Nova Herculis.*
(*a*) *Hα*
(*b*) *Hβ*
(*c*) [*OI*] 6364, 6300, 5577 (*mean*).
(*d*) [*NII*] 5755
(*e*) *FeII* a^4G–z^4F.
(*f*) *FeII* a^6S–z^6P.

The difference between the two curves of iron lines belonging to different multiplets is striking. In the multiplet *aG–zF* containing the lines λλ5316, 5276, 5234, 5197 the plot of mean *R/V* values against magnitude in Fig. 3*e* indicates that all along the observations fit more closely a straight line with a slope less steep than in the case of part (2) in the curve of the former multiplet, as shown in Fig. 3*f*.

In the case of the [*OI*] lines λλ6364, 6300, 5577 the result is different altogether. In Fig. 3*c* the mean of the ratios of these three lines is plotted against magnitude. *R/V* values of the line λ6364 may have been too large in February mainly because of its possible blending with *Si*II λ6371·36. The ratio *R/V* for the line λ5755 of forbidden ionized nitrogen plotted against magnitude is given in Fig. 3*d*, where it can be seen that the slope is less steep than in the case of [*OI*] lines, which show the most rapid decline of *R/V* values with the nova's decrease in brightness. The rates of decline of the ratios *R/V* for a decrease in brightness of 1 mag. are given in Table III.

To return to Fig. 2 where, in the case of *Hz*, the curves are not completed, there were not enough observations due to the lines being over-exposed on most of the available plates. For the same reason λλ6364, 6300 were not studied in

TABLE III

Rates of decline of R/V per unit magnitude decline

Lines	Rates	Lines	Rates
Hα	(1) −0·15 (2) +0·15	λ 5755	+0·15
Hβ	−0·15		
λ 6364, λ 6300, λ 5577 mean	+0·17	λ 5316, λ 5275, λ 5234, λ 5197 mean	+0·03
λ 6300, λ 5577 mean	+0·16	λ 5169, λ 5018, λ 4924 mean	(2) +0·09 (1) −0·20

(1) *and* (2) *refer to portions of curves in Fig.* 3*a and* 3*b from left to right.*

the early stages of the period. λ5755 on the other hand was too faint at the beginning to be detected photometrically. From January 16 up to the end of 1935 January the four ionized iron lines λλ5316, 5275, 5234, 5197 were crossed by absorptions or blended with other emissions, so that the values of the maxima could not be deduced with certainty. In Fig. 2 the corresponding curves are partly represented by broken lines to indicate that they are not continuous. Such unreliable values are shown as open circles in the diagrams.

6. *Notes on diagrams.*—It is observed in the diagrams that there are no great variations in the *R/V* curves before about January 20 except in the case of *Hβ*. The light curve also does not exhibit large fluctuations between January 1 and January 15. In the following we shall summarize the observed behaviour of the various lines and multiplets and compare whenever possible the actual intensities with those given by F. L. Whipple and C. Payne-Gaposchkin.*

Hydrogen lines: It could be said generally that the hydrogen lines *Hα* and *Hβ* did not follow exactly the same course, though data on *Hα* are by no means sufficient to correlate its behaviour with that of *Hβ*. We are not primarily discussing here the structure of the lines, their width, or profiles, but the comparison of the relative intensities of their components. January 13 was the first date when *Hα* started to show two peaks; for quite a long time early in February a double-peaked top was clearly visible, as indicated in Fig. 4, except on the dates when the line was too strongly exposed to provide good measures. The same double-peaked top was markedly shown by *Hβ* at the time when it was exhibited by *Hα* but was less clearly observed at a later date. The red edges of *Hα* were always better defined than the violet ones, but in *Hβ* both of the two edges were generally indistinct except for a short time at the end of 1935 December, early in January (Fig. 4) and again after the end of 1935 January. In the case of *Hβ*, particularly where we have sufficient observations, intensities relative to the continuous background increase strongly after the time of the nova maximum, for a few days early in February and again in March. This is not so noticeable in the case of iron lines which remain more or less constant in brightness till the end of January and afterwards, after some fluctuations, exhibit a slight decline. *Hα* on the other hand showed a considerable rise from the end of January till about February 5. From January 13 till February 9 the violet was stronger than the red and from this date for about a week the red was stronger. No observations are available from February 15 to the end of March,

* *H.A.C.*, No. 414, 1937.

the red getting much stronger again until early in April when the process was reversed. In $H\beta$ the violet component was much stronger during the first half of January, then, earlier than in the case of $H\alpha$, it got fainter than the red to become stronger again from about February 20 to March 5, during which time observations of $H\alpha$ were lacking. From March 5 to the end of the period under observation on April 3 the violet component was fainter, on the whole behaving exactly opposite to $H\alpha$. Observations during this last period were doubtful, however, owing to the presence of [*Fe*II] λ4874 which affected the red component.

The rise in the intensity of $H\beta$ during the middle of January and its fall in February, as seen in Fig. 1, agrees with that given in the Fig. 1 of *H.A.C.* The maximum in the next rise about March 5 is also indicated in the Harvard observations of the same date. Two peaks shown in our curves for February 5 and March 22 are not visible in the *H.A.C.* curves.

Oxygen lines [*O*I]: The forbidden oxygen lines, in particular $\lambda\lambda$6364, 6300, are greatly strengthened, relative to all other lines except $H\alpha$. At the time of the deep light minimum all the lines of the oxygen multiplet have been greatly strengthened.

(*a*) [*O*I] λ6300: The violet component is stronger than the red from January 22 till February 6, in agreement with the drop in the light curve, and weaker for the following four days when the brightness increases. From February 10 to 17 the violet component is stronger at the time of decreasing light. When the light increases, but a little later, the red component becomes stronger. The violet component strengthens for the third time, coinciding, with a short lag, with the drop in the nova brightness. With the rise in the light curve on February 25 the intensity of the red rises abruptly, relative to the violet, to fade out together with the gradual fading of the nova. The close correlation between the strengthening of the violet relative to the red and the fall in the light curve, is shown also in the subsidiary fall on March 20 just before the collapse of the nova.

(*b*) [*O*I] λ5577: This line followed exactly the behaviour of λ6300, but there were not enough observations to show the final disappearance of the red component.

(*c*) [*O*I] λ6364: The R/V curve differs at the beginning of the period under investigation from that of the two previous oxygen lines. This effect might be due to interference by the line *Si*II λ6371, as was mentioned before. These observed deviations are not marked at all from the end of February till the end of the period of observation.

The intensity of the mean of $\lambda\lambda$6364, 6300 in *H.A.C.*, Fig. 2, is low early in January, rises at the end of January and attains a maximum on about March 4, which is present on our diagrams of Fig. 1 in the case of λ6300 and λ5577, but not in that of λ6364. From the middle of March onwards intensities of $\lambda\lambda$6364, 6300 increase steadily, while λ5577 fades gradually in the *H.A.C.*, Fig. 2. This is somewhat different from the rise shown in our diagram for all red components of $\lambda\lambda$6364, 6300; there is also no confirmation in our diagram of the gradual falling of λ5577 found by the Harvard observers.

[*N*II] λ5755: The intensity of this line, which is weak in the early stages of the nova, shows pronounced strengthening from the beginning of March onwards. The R/V curves do not show the same behaviour as in the case of the [*O*I] lines,

In one instance on February 18 when the ratio R/V is a maximum for λλ6300, 5577, the nitrogen line shows a minimum and the same happens again on March 24. On the other hand, when the red component of the nitrogen line is strengthened, all three lines of forbidden oxygen showed a strengthening of the violet component relative to the red one. Finally the red component weakens considerably relative to the violet without vanishing completely, as in the case of the oxygen lines.

*Ca*II (K) 3933: This line is near the end of the spectral region in focus on the Cambridge plates, and therefore there are rather few observations, and the determination of the intensity level of the continuous background is uncertain in some cases. The wide absorption affects the violet peak and in addition weakens it, especially at first.

E. G. Williams *, in his valuable work on the intensities of bands in the spectrum of Nova Herculis, has given amongst others the contours of the K line for some selected dates. The values that could be deduced from his diagram for the intensities of the two peaks are in fairly good agreement with the R/V ratios derived in the present work.

From February 5 to March 4 the red component varies strongly approaching the violet one in the R/V curve. After a small increase lasting about ten days the R/V curve gradually declined.

The curve shown in Fig. 1 is similar to the corresponding curve in *H.A.C.*, Fig. 2, both in its steadiness from February 5 to 15 and in its falling and rising in the beginning of March. The drop at the end of February, according to Harvard observations, is not visible on our diagram for lack of observations. The *Ca*II (H) line could not be measured and compared with the (K) line. The individual lines of the following two iron multiplets are not included in the diagrams.

*Fe*II *lines* a^4G–z^4F^0: With the exception of the line λ5275 the three other lines of this multiplet show a gradual rise in the intensity of the red component with respect to the violet one early in January, with a maximum on the 10th in the case of λλ5316, 5234 and on the 13th in the case of λ5197. The doubtful observation of λ5275 on January 13 is probably the reason for that different behaviour. The ratio R/V is nearly always greater than unity until about the middle of March, with the exception of the line λ5197. The two outstanding maxima in the R/V curves for this multiplet are on February 5 and 18, most conspicuous in the case of λλ5316, 5275. These dates correspond to the end of the first secondary light minimum in February and the middle of the second secondary minimum late in February respectively. Other maxima in the case of some lines are on February 23, March 4, 21 and 25; the second of which corresponds to a light maximum and the third falls in the middle of the last subsidiary minimum. Finally we notice that the complete disappearance of the red component is much the same as the forbidden lines of oxygen. The two components fade away gradually at the time of the great minimum of the nova, the red more quickly than the violet. The R/V curve for the mean of all lines of this multiplet in Fig. 2 indicates that the R/V values were altogether greater than 1 until March 5 with shallow variations the maxima of which coincide except the last one with high minima. In contrast to the oxygen lines for which the R/V curves are more or less parallel to the light curves, the same curves show a reverse behaviour in the case of *Fe*II lines.

* *M.N.*, **99**, 91, 1938.

*Fe*II a^6S–z^6P^0: In the case of all three lines of this multiplet there is a considerable increase in the intensity of the red relative to that of the violet components at the end of the second minimum early in February. On the other hand, there is a minimum clearly visible in the R/V curves of the two lines $\lambda\lambda$ 5018, 4924 and to some extent in that of λ 5169 from February 24 to 26, that is near the time of the light minima on February 23. Again there is a marked minimum in the R/V curves of all three lines on March 22 at the time of the last subsidiary minimum. On the other hand, there are two maxima on February 5 and 28, well marked in the case of $\lambda\lambda$ 5018, 4924, at the time of the light maxima of the nova.

As regards total intensity, Fig. 1 shows that this attains a maximum on February 5, 17, March 5 and 22, while in the *H.A.C.*, Fig. 2, the maxima are on January 29, February 10, 22, March 4 and 22. Our minima are on February 3, 12 and 26, while the Harvard observers find them for February 5, 12, 25, March 18 and 27. There appears to be fair agreement between the Harvard and our own intensities and their variations, except that we do not confirm the Harvard maximum on January 29.

The curve for the mean of the whole multiplet in Fig. 2 starts with very low values of R/V, rising gradually to the maximum on February 5. The subsequent variations show no clear correlation with the light variations of the nova. The R/V values decrease gradually before and at the time of the nova collapse.

*Fe*II λ 5535 b^2H–2^4F^0: This line starts with a ratio $R/V<1$, which gradually increases during January, reaching a maximum on February 11 at the time of a small minimum in the light curve. A deep minimum, rather incompletely observed, on February 20 is found roughly on the downward slope of a deep minimum in the light curve on February 23. Although observations are not sufficient, a striking feature is the general rise of the R/V curve for this line in contrast to those for all other lines. A general fall follows this rise.

*Fe*II λ 4233 b^4P–2^4D^0: In this line the red component is generally stronger than the violet one all through the month of February, also during the two secondary minima at the beginning of February and the end of March. Here it is in agreement with the behaviour of the oxygen lines. The maxima are not shown clearly enough to show with which of the two iron multiplets it agrees.

Discussion

The photometric measures of the emission lines appear to point to the following conclusions: (*a*) Lines of different elements, namely those of hydrogen, forbidden oxygen, ionized iron and calcium and forbidden ionized nitrogen, behaved differently. (*b*) The same was true of lines of the same element, but of different energy levels as noticed in ionized iron lines. Minor variations appear to be present in the behaviour of lines of the same multiplets. (*c*) Particularly striking was the correlation between successive strengthening of red and violet components of the forbidden oxygen lines and the fluctuations in the light curve of the nova.

$H\alpha$ and $H\beta$ were both exceptional in the regular rate of strengthening of the red with respect to the violet components and their correlation with the decrease of brightness of the nova. There might be a reverse tendency with the decrease of brightness of the nova at its later stages, but there were not sufficient observations to show this definitely. Oxygen lines are of greatest interest in

FIG. 4.—*Microphotometer tracings of double emission lines in Nova Herculis.*

Diagram	*Plate*	*Date*	*Lines (from left to right)*
(1)	T 92	1935 *March* 21	[OI] 5577
(2)	G 1259	1935 *April* 2	[OI] 6300, 6364
(3)	G 1186	1934 *December* 31	*Hβ* 4861
(4)	G 1213	1935 *February* 11	*Hα* 6563
(5)	G 1249	1935 *March* 20	*Fe*II 5018
(6)	G 1249	1935 *March* 20	*Fe*II 5169, 5197, 5234, 5275, 5316
(7)	G 1183	1934 *December* 30	*Fe*II 5535
(8)	T 390	1935 *March* 5	*Ca*II 3933
(9)	T 92	1935 *March* 21	[*N*II] 5680, [*N*II] 5755
(10)	T 387	1935 *March* 4	*Fe*II 4233
(11)	G 1249	1935 *March* 20	*Fe*II 4924

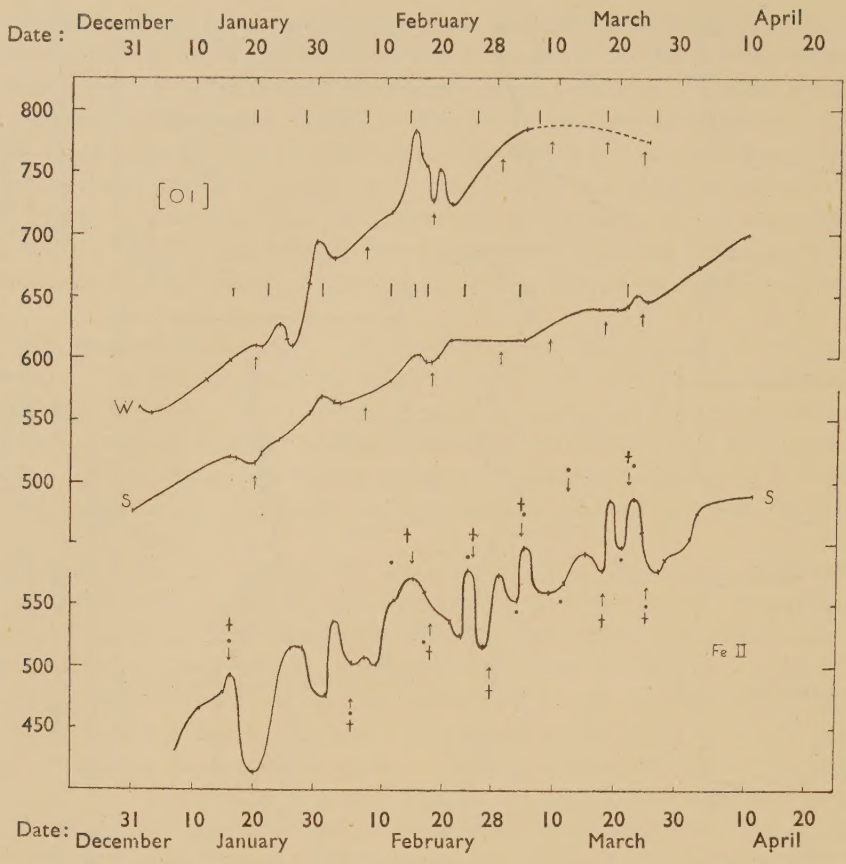


FIG. 5.—*Upper portion*: *Separation* S *and width* W *in km/sec. for* [OI] (*Merrill*) *plotted against time. Arrows indicate times of maxima R/V for* [OI] *lines. Vertical lines also indicate times of maxima* (*above*) *and minima* (*below*) *in the light curve.*

Lower portion: *Separation* S *in km./sec. for Fe*II *lines* (*Stratton*). *Times of maxima and minima R/V* (*Fe* II) *are indicated by upward and downward arrows respectively* (*for* $Fe\,a^4G\text{–}z^4F^0$, *and for* $a^6S\text{–}z^6P^0$).

Note the coincidence of maxima R/V with minima S (*and vice versa*) *for Fe*II *only.*

this investigation, since there is practically no absorption in their neighbourhood, and very little interference with other lines. At the top of Fig. 5 the measures of width (W) and separation (S) of the two components, as given by P. W. Merrill *, are shown for these oxygen lines. It is noted that the separation increases practically uniformly. The slope rises gradually while the width increases particularly strongly on January 30 and February 15, that is at the downward slope of two minima in the light curve. This means that there might be some emission bands from new expanding shells superimposed on the emission lines, which become strengthened during these particular times when the light curve is falling from maxima to minima. The two components of these lines are otherwise quite undisturbed. The times when the R/V values were greatest are indicated in the figure; they do not appear to be related to anything in the W and S curves. These maxima repeat at regular intervals, as seen in Table IV, every nine days. On nearly the same dates they correspond to points of maxima in the light curve. The values of these R/V maxima are of the same order of

* *Ap. J.*, **82**, 413, 1935.

magnitude. The curves rise rather more steeply to maxima, then they fall to the consecutive minima. There is a slight difference between the lines λλ6300, 5577 which belong to different multiplets.

At the bottom of Fig. 5 the mean separation of the two components for several iron lines is also given for the whole period, as kindly communicated to the author by Professor Stratton.* Apart from the general rise in the curve of S, certain fluctuations are worth considering. From the arrows pointing to dates of maxima and minima of R/V values, it is easily seen that times of maximum separation coincide with minimum values of R/V and vice versa. Nothing of this sort is noticeable in the case of oxygen lines. Points for curve S were taken from the means of several iron lines. The measures of percentage depths of *Fe*II absorption lines, as given in Table III of Williams's paper † show maximum values when the separations S given in Fig. 5 are least and *vice versa*. This can be explained by the assumption that absorption "eating into" the violet components increases the R/V values, diminishing the measured separations and *vice versa*. It seems possible to attribute to this factor most of the observed variations of the ratio R/V for ionized iron lines. The differences between the different iron multiplets have been pointed out before, but it can be said that both multiplets separately show maxima and minima of R/V, which account satisfactorily for minor anomalies in the radial velocity measures. For hydrogen, nitrogen and calcium this comparison could not be made for lack of observations.

Turning back to the periodic variations of the R/V values for the oxygen lines and the corresponding variations in the light curve, one might point to a paper by P. M. Ryves ‡ in which he suggests the existence of periodic changes in the light curve with periods of 6·2 days between about 1934 December 25 and the end of 1935 January, and of 4 days between about 1935 January 25 and March 3. Table IV gives the dates of maxima and minima and the number of days between each. The lengths of periods in the light curve are rather unequal, pointing to a sort of alternate long and short periods with corresponding deep and shallow minima. The average length of these periods is 9·3 days. On the other hand the R/V curves show fairly regular periodicities, most clearly in the case of [*O*I] lines where their presence cannot be attributed to errors of observation.

The average length of this period in the R/V curves taken from λλ6300, 5577 is about 9 days. The period suggested earlier for the light curve is about twice as long as Ryves's. This may mean that either the secondary periodic variations in the R/V curves are of the order of observational errors or that Ryves's light curve should be smoothed further when it might show the same period of about nine days.

Assuming that there are real periodic variations in the R/V curves we have to find their origin. The existence of the red and violet components is probably due to the ejection of two main jets of matter from the star. To explain those variations one might think of three possibilities: (*a*) The ejection consists in a kind of pulsation affecting the red and violet components simultaneously. (*b*) The ejection is continuous, but is interrupted at intervals by some obscuring matter appearing between the central star and the condensations of ejected matter.

* Unpublished.

† E. G. Williams, *M.N.*, **99**, 91, 1938.

‡ P. M. Ryves, *M.N.*, **95**, 535, 1935.

TABLE IV

Dates of maxima and minima in light and R/V curves for the emission lines

Maxima

	January		February			March				Average period
Light curve	20	28	7	14	26	...	7	18	26	9·3
Hβ	...	...	5	...	...	...	...	20	...	
Mean all [*O*I]	20	...	7	18	28	...	9	...	24	
Mean [*O*I] λλ 6300, 5577	20	...	7	18	28	...	9	18	24	9·0
Mean *Fe*II a^4G–z^4F^0	...	...	5	17	...	...	11	21	25	
Mean *Fe*II a^6S–z^6P^0	...	...	5	18	28	5	...	18	25	
All *Fe*II	...	...	5	18	28	...	...	18	25	
*Ca*II (*K*)	...	...	5	18	...	...	...	...	...	
[*N*II] λ 5755		...	5	...	27	...	...	...	26	

Minima

	January				February			March		
Light curve	16–17	21–25	31	4	9–13	17	23	4	...	21–22
Hβ	13	...	...	...	...	17	26	...	11	26
Mean all [*O*I]	10	...	...	3	15	...	26	4	...	20
Mean *Fe*II a^4G–z^4F^0	16	...	...	...	12	...	24	5	...	23
Mean *Fe*II a^6S–z^6P^0	16	...	...	...	14	...	25	4*	12	22
All *Fe*II	16	...	...	4	...	...	...	...	...	...
*Ca*II (*K*)	...	...	...	4	...	...	...	...	...	...
[*N*II] λ 5755	...	...	...	...	17	...	25	5	19	24

* This value is obtained from one line only.

(*c*) As in Nova Persei, according to Oort *, collision between the ejected matter and some existing outer layers may increase the excitation and lead to maxima in one of the emission components. The hypothesis of a rotation of the whole system would explain the effects also, but there appears to be no evidence in the velocity measures for such a rotation.

It is likely that in addition to main ejections others of a minor character took place in other directions. Recent photographic observations show two large masses on both sides of the faint central star along the major axis and two smaller condensations at right-angles to them. Spectroscopic evidence of such knots is found on the Mount Wilson spectrograms, which show them to be well inside the outermost surface of the envelopes. Stratification is suggested by these results and may well explain the different behaviour of different lines, e.g. H, [*O*I], *Fe*II and [*N*II].

Spectra of the period between the collapse of the nova and the time of the more recent observations in 1942 would be of particular interest in connection with this problem.

I have to express my deepest acknowledgment to the continuous help and exceedingly valuable suggestions of Professor Stratton, who also proposed this problem to me for study, and to Dr Brück for supervising and invaluable criticism.

Solar Physics Observatory,
Cambridge :
1947 *March* 6.

* Unpublished.